THE STRUCTURE AND STATISTICS OF LANGUAGE JOINTLY SHAPE CROSS-FREQUENCY DYNAMICS DURING SPOKEN LANGUAGE COMPREHENSION

PREPRINT VERSION

Hugo Weissbart*
Donders Centre for Cognitive Neuroimaging
Donders Institute
Nijmegen, The Netherlands
*hugo.weissbart@donders.ru.nl

Andrea E. Martin
Donders Centre for Cognitive Neuroimaging
Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands
andrea.martin@mpi.nl

September 21, 2023

ABSTRACT

Humans excel at extracting structurally-determined meaning from speech, despite the inherent physical variability of spoken language (e.g., background noise, speaker variability, accents). One way to achieve such perceptual robustness is for the brain to predict its sensory input and, to some extent, the linguistic content, based on its internal states. However, the combinatorial nature of language, which on one hand endows language with its unboundedness and expressive power, also renders prediction over a sequence of words a non-trivial and, at the very least, non-Markovian affair. How neural infrastructure allows for linguistic structures, e.g., the hierarchical organisation of phrases, to be jointly processed with ongoing predictions over incoming input is not yet well understood. To wit, this study takes a novel perspective on the relationship between structural and statistical knowledge of language in brain dynamics by focusing on phase and amplitude modulation. Syntactic features derived from constituent hierarchies, and surface statistics based on word sequential predictability obtained from a pretrained transformer model, were jointly used to reconstruct the neural oscillatory dynamics during naturalistic audiobook listening. We modelled the brain response to structured and statistical information via forward encoding models, and found that both types of features improve decoding performance on unseen data. Results indicated a substantial overlap between brain activity involved in both types of information, suggesting that the classic viewpoint that linguistic structures and statistics about them can be separated as a false dichotomy when language is processed in the brain. Syntactic features aided neural signal reconstruction over a longer period of time; in contrast, the effect of statistical features is comparatively shorter, but is tightly bound to the phase of neural dynamics, suggesting involvement in the temporal prediction and alignment of cortical oscillations involved in speech processing. Both features are jointly processed and contribute to ongoing neural dynamics during spoken language comprehension, and are locally integrated through cross-frequency coupling mechanisms.

Keywords MEG, spoken language comprehension, surprisal, syntactic structure, phase-amplitude coupling, cross-frequency coupling, forward encoding model

*corresponding author

1 Introduction

Humans comprehend language despite the high variability in the physicality of acoustic and phonetic inputs, as in noisy environments or when encountering new speakers or accents. Such robustness may stem from the projection of stored knowledge via anticipatory mechanisms; predictive processing is a widespread cortical computational principle, or more generally, a framework in which predictions play an active role in the processing of upcoming information streams(e.g., Keller and Mrsic-Flogel, 2018). The incoming information may be sensory or solely endogenous neuronal activity from other brain regions, and the output of such process may be perceptual, motor, or cognitive (Bastiaansen and Hagoort, 2006; Friston, 2010; Luc H. Arnal and Giraud, 2012; Giraud and Luc H Arnal, 2018). Although this framework originates from the field of perception (Rao and Ballard, 1999), evidence of such processing during language comprehension is also abundant (e.g., Keitel et al., 2016a; Zoefel, Oever, and Alexander T. Sack, 2018; Di Liberto, Crosse, and Lalor, 2018; Keith B. Doelling et al., 2019; Heilbron et al., 2022; Andrea E Martin, Monahan, and Samuel, 2017; Andrea E Martin, 2016; Andrea E. Martin, 2020; Mante S Nieuwland et al., 2018; Mante S Nieuwland, 2019; Mante S. Nieuwland et al., 2020). However, language comprehension mainly occurs when inputs are not fundamentally predictable (as a new message, encoding the intended meaning from the speaker, must be comprehended; Andrea E Martin, 2016). Arguably, the relay of novel information often depends on sequences that do not have a high likelihood of occurrence. However, there are different levels of granularity, each benefiting from prediction from other representations, such as phonemes predicted from preceding items and spectro-temporal characteristics of the sound segment (Andrea E Martin, Monahan, and Samuel, 2017; Kaufeld, Naumann, et al., 2020). Furthermore, an essential feature of human language is its display of nested syntactic structures over which meanings are computed. Traditionally, language's inherent unboundedness and generative aspect have often been seen as being in putative opposition to distributional and statistical accounts of language processing (e.g., Chomsky, 2013; Andrea E Martin and Doumas, 2017). In contrast to this view, in the present study, we synthesise these positions and present a framework wherein the syntactic structure and statistical cues are jointly processed during comprehension (see Andrea E Martin, 2016; Andrea E. Martin, 2020). The phase of neural signals is a candidate for carrying information both about statistical and structural features (Ten Oever and Andrea E Martin, 2021; Jonathan R Brennan and Andrea E Martin, 2020; Jensen and Lisman, 2000; Lisman and Jensen, 2013; Ten Oever and Alexander T Sack, 2015; Ten Oever, Meierdierks, et al., 2020). We further investigate how the phase of low-frequency oscillations couples with power at higher frequencies with discriminatory spatio-temporal signatures for either statistical or structural cues.

To operationalise syntactic processing, we constructed a set of word-level features which describe critical characteristics of the underlying constituency parse or tree structure. Those syntactic features are thus derived from hierarchical binary trees operating at the sentence level and are de-lexicalised. They represent aspects of syntactic trees (viz., number of brackets, depth in the tree) which by themselves can elicit a similar simulated response reflecting tracking of syntactic structures as in Ding, Melloni, et al. (2016) and Frank and Yang (2018). These syntactic metrics have been used to study the effect of syntactic operations such as unification or integration ("merge" operation in the minimalist program) or integration of an item into a larger structure or the depth, a proxy for ongoing complexity, of the syntactic tree at a given word (Bastiaansen, Magyari, and Hagoort, 2010; Nelson et al., 2017; Jonathan R. Brennan and Hale, 2019; Hale et al., 2022).

Surface statistics on the other hand are estimated from the bare sequence of words with no information whatsoever beyond the sequence itself (J. Elman, 1990; Frank and Yang, 2018). However, this does not necessarily imply such statistics will not carry any information about hierarchical structures or about syntax, a topic that is currently under debate (Slaats and Andrea E Martin, 2023; McCoy et al., 2023; Mahowald et al., 2023; Levy, 2008; Van Schijndel and Linzen, 2021). What is indisputable, however, is that such statistical metrics are built from the sequence of words and optimised to learn a probability distribution conditioned on the sequence context of preceding words. To learn this distribution, words are presented to a system, such as a recurrent neural network, and the system's task is to predict the next word based on the series of preceding words encountered so far (this memory is embedded in the model's architecture). Non-linear models, like Long Short-Term Memory network or Transformers, can manipulate information such that they could potentially encode a form of second-order structured information within their memory (hidden states, latent representations). From an information-theoretic perspective, surface statistics can give an estimate of the self-information (viz., surprisal) and uncertainty (viz., entropy) measured at the word level while conditioning on the observed context of previous words. Importantly we note that these models, despite their increasingly complex architecture, do not have rule-based knowledge (structurally-)embedded in their learning algorithm and solely make predictions from an estimate of the conditional probability distribution over the vocabulary that is updated through their inner (recurrent) dynamics. However, we note that recent models, such as larger GPTx, appear to perform better on a series of syntactic paradigms (Mahowald et al., 2023; McCoy et al., 2023) compared to older architectures that were trained jointly with grammatical rules and structural information along with the word sequence (such as RNNG, Grammatical recurrent neural network; but see(Hale et al., 2022; Antonello and Huth, 2022; Jain et al., 2023; Guest and Andrea E Martin, 2023)).

Cortical activity is modulated by both surprisal and entropy of words during language comprehension (e.g., Weissbart, Kandylaki, and Reichenbach, 2020; Nelson et al., 2017; Frank, Otten, et al., 2015; Frank and Willems, 2017). It has also been proposed that the ongoing phase of neural signals may be modulated by the predictability of words, thus affecting the temporal prediction of syllable and word onset (e.g., Ten Oever and Andrea E Martin, 2021; Ten Oever and Alexander T Sack, 2015). In the present study, we investigate cortical dynamics in phase and amplitude in response to variation in both statistical and syntactic features.

In light of this apparent dichotomy, and in the context of the classic debate in cognitive science regarding the role of statistical information in language representation and processing (Saffran, Aslin, and Newport, 1996; Marcus, 1998; Chomsky, 2014; MacDonald, Pearlmutter, and Seidenberg, 1994; J. L. Elman, 1991), we then ask whether *distinct sources and temporal dynamics in response to structure or statistics of the linguistics input.* And also to *what extent does their individual contribution explain neuroimaging data, is the whole better than the sum of its parts?* We hypothesise that they jointly contribute to explaining variance in the MEG data while presenting overlapping spatiotemporal sources. However, the dynamics might disentangle them further as the predictions and statistical inference seem to be a widespread phenomenon in cortical computation, the organisation of linguistic units into nested hierarchical structures, at least at first blush, may be related to hierarchical processing in other domains in some ways, but not others (Bellmund et al., 2018; Boorman, Sweigart, and Park, 2021; Coopmans, De Hoop, Kaushik, et al., 2022; Coopmans, Kaushik, and Andrea E Martin, 2023; Coopmans, De Hoop, Hagoort, et al., 2022; Ten Oever, Carta, et al., 2022). We thus further hypothesise that brain responses to statistical and structural features are orchestrated synchronously but with distinct time scales. That differentiation arises as information flows from one system of prediction to, potentially, another, which would derive and compute different combinatorial aspects of structure building and semantic processing.

To tap into the orchestration of spatio-temporal dynamics, we measured brain activity with magnetoencephalography (MEG), allowing for a time-resolved recording of neural activity. The temporal resolution of MEG data is fine enough to measure power modulation in a wide range of (high) frequencies together with the phase of a slower frequency range. The role of cortical oscillations in neural computation is still unclear; nonetheless, an increasing number of studies have now attributed some functional role to different frequency bands, not only for low-level perceptual or sensory processing but also in relation to speech processing. Delta and theta rhythms (1-4Hz and 4-8Hz respectively) play an important role in the neural tracking of the acoustic envelope, for instance (Peelle and Davis, 2012; Di Liberto, Crosse, and Lalor, 2018; Di Liberto, O'Sullivan, et al., 2015). It has been hypothesised that low-frequency oscillations rhythmically modulate neuronal excitability to match the rhythm and landmarks of the acoustic stream, perhaps reflecting speech segmentation mechanism (Giraud and Poeppel, 2012; Keitel et al., 2016b; Oganian, Kojima, Breska, Cai, Findlay, E. Chang, et al., 2023; Keith B Doelling et al., 2014) but also facilitating the processing of syntactic information via synchronisation to lower level acoustic cues (Bastiaansen and Hagoort, 2006; Coopmans, De Hoop, Hagoort, et al., 2022; Kaufeld, Bosker, et al., 2020; L. Meyer and Gumbert, 2018a; Kösem, Wassenhove, and Wassenhove, 2017; Ten Oever, Carta, et al., 2022; Slaats, Weissbart, et al., 2023; Zioga et al., 2023). Moreover, it has been shown that those low-frequency oscillations couple their phase with power of higher frequency broadband power (phase-amplitude coupling, PAC thereafter) and such coupling has an impact on behavioural response (Lakatos et al., 2008; Luc H. Arnal, Keith B. Doelling, and Poeppel, 2015).

We further investigate the relation of word level computation at different timescales, following the study from Peter W. Donhauser and Baillet (2020). Indeed, they observed distinct role during the prediction of phoneme sequences for theta and delta rhythms. With theta treated as a read-out of the sensory sampling mechanism bound to maximize the expected information gain. On the other hand, delta is encoding non-redundant information in their study. That is, novel information deviating from internally generated predictions, which must *update* the internal model (Ten Oever and Andrea E Martin, 2021; Andrea E. Martin, 2020; L. Meyer, Sun, and Andrea E Martin, 2020a; L. Meyer, Sun, and Andrea E Martin, 2020b). Top-down predictions, but also updates, have been already linked to beta power modulation. All in all, this mechanism, if supported via delta-phase, is bound to endogenously generated predictions and thus to the internal model of the listener. Thus, in line with Peter W. Donhauser and Baillet (2020), we expect a stronger delta-beta coupling for non-redundant information. This is approximated by the *surprisal* feature of the statistical model.

However, it is still widely debated whether measured oscillatory activity in response to speech stimulus truly comes from an oscillatory mechanism or is a by-product of measuring a response to a (pseudo-)rhythmic stimulus, that is, a series of evoked responses (see Andrea E Martin and Doumas, 2017; L. Meyer, Sun, and Andrea E Martin, 2020b; L. Meyer, Sun, and Andrea E Martin, 2020a; Ten Oever, Kaushik, and Andrea E Martin, 2022; Haegens and Golumbic, 2018; Obleser and Kayser, 2019b). The interpretation of the role, cause and effect, of low-frequency oscillations in particular is not clear. While some studies claim that there is evidence that neural tracking to speech envelope is solely due to evoked response convolved with acoustic edges (Zou et al., 2021; Oganian, Kojima, Breska, Cai, Findlay, E. F. Chang, et al., 2023), other studies put forward evidence for an entrained endogenous oscillator which shows sustained activity and phase-related behavioural modulation (Keith B. Doelling et al., 2019; Teng et al., 2018; Zoefel, Oever, and Alexander T. Sack, 2018; Ten Oever, Kaushik, and Andrea E Martin, 2022).

In the present study, we ask whether syntactic and statistical word-level features provide complementary cues to the processing of speech input. Moreover, we investigate how putative cortical oscillations are distinctively modulated by linguistic features. The aim is to discuss on a higher level how different frequency bands orchestrate the processing of linguistic units. Those units are favoured differently, whether from predictions on sequential statistics, a potentially domain-general feat, or mirror integration into larger linguistic compounds, thus reflecting key, putatively language-specialised computations.

2 Methods

A total of 24 participants (18 women, between 18 and 58 years old) completed the experiment. All participants were right-handed native Dutch speakers with no reported fluency in French despite incidental exposure. Participants self-reported their (in)ability to understand a sentence in French. The study was approved by the ethical commission for human research in Arnhem and Nijmegen (CMO2014/288). Participants were given monetary reimbursement for their participation.

2.1 Experimental design

Participants were asked to listen and pay attention to several audio stories while we simultaneously recorded their MEG. The stimulus consisted of Dutch short stories (from Hans Christian Andersen and the Brothers Grimm) and French stories (from Grimm, E. A. Poe and Hans C. Andersen), totalling 49min and 21min, respectively. All stories were divided into short story parts lasting between 5 and 7 minutes (leaving a total of 9 Dutch story parts and 4 French ones). Each story part was presented without interruption, while participants fixed a cross in the centre of the presentation screen. Participants were prompted on-screen with five multiple-choice questions between each part to assess comprehension. Stimuli were presented using the Psychtoolbox library on Matlab (D. H. Brainard, 1997; Kleiner, D. Brainard, and Pelli, 2007).

MEG data were acquired at 1200Hz using a CTF 275-channel whole-head system (VSM MedTech, Coquitlam, Canada) in a magnetically shielded room. The MEG system was equipped with 275 first-order axial gradiometers with a baseline of 5 cm. The head position was measured before and after the experiment using five head position indicators (HPI) coils. The HPI coils were activated every 200ms during the experiment to monitor head movements. The head position was corrected for each story part using MaxFilter software (Elekta Neuromag, Helsinki, Finland). Participants could self-pace the start of a trial after answering the behavioural comprehension question, allowing them to pause between stories blocks. Each participant also had a structural MRI scan (T1-weighted) using a 3T MAGNETOM Skyra scanner (Siemens Healthcare, Erlangen, Germany). The MRI scan was used to reconstruct the cortical surface of each participant using FreeSurfer software (Martinos Center for Biomedical Imaging, Charlestown, MA, USA). The cortical surface was used to project the MEG sensor data onto the cortical surface using a linearly constrained minimum variance beamformer (Van Veen et al., 1997). Finally, we measured the head shape of each participant using a Polhemus Isotrak system (Polhemus Inc., Colchester, VT, USA) to co-register the MEG and MRI data.

2.2 MEG preprocessing

The original MEG data were recorded at 1200Hz. We first resampled the data at 200Hz after applying and anti-aliasing low-pass filter to the data. Noisy channels and flat channels were marked as bad for interpolation and to be discarded in subsequent analysis (computation of covariance or ICA algorithm). We removed blink artefacts by matching ICA component time courses to measured EOG and similarly removed heartbeat artefacts. ICA decomposition was run on band-passed data: we applied a band-pass filter between 1 and 40Hz to the data. Finally, we applied a notch filter at 50Hz to remove line noise.

We then used all data to compute a data covariance matrix, and used the covariance matrix to compute a noise-normalized linearly constrained minimum variance (LCMV) beamformer (Van Veen et al., 1997). The LCMV beamformer was created using a 7 mm grid with 3 mm spacing.

Time-frequency Analysis was performed on the MEG data using the MNE software package (Gramfort et al., 2013). We first epoched the data around word onset, with no baseline applied. We then used the Morlet wavelet transform to compute the time-frequency decomposition of the MEG data. The Morlet wavelet transform was computed varying number of cycles from 2 to 7 cycles. We used a logarithmically spaced grid of frequencies between 3 and 80 Hz, with 32 steps. The time-frequency decomposition was computed for each epoch. We extracted the average power by taking the absolute of the complex Fourier coefficients and the inter-words phase constituency by summing the complex Fourier coefficients after normalising them by their absolute value.

2.3 Linear Models

Most of the analyses performed relied on linear models. We computed forward encoding models which map stimulus features to MEG data. Such models are also called temporal response functions (TRF) (Crosse et al., 2016; Obleser and

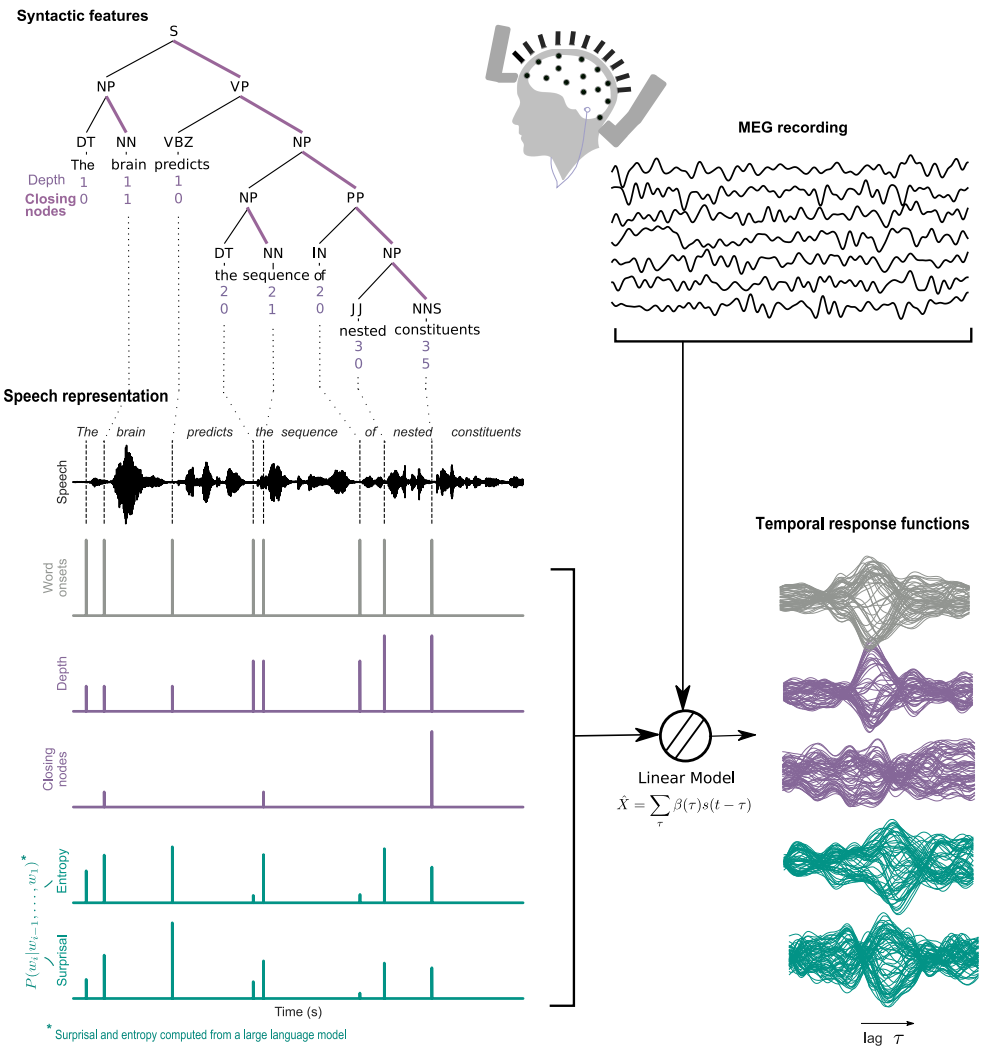


Figure 1: Example of features used. The syntactic features (A), i.e. tree depth and the number of closing nodes are shown below each constituency tree leaves. The MEG recordings (B) together with the stimulus representation (C) are used to compute temporal response functions (D).

Kayser, 2019a). This approach gives us a way to assess the importance of each feature in the stimulus in explaining the MEG data. The obtained model coefficients are also directly interpretable in terms of modulated neural activity, as opposed to filters learned from backward models (Haufe et al., 2014). In this section, we will first present how each stimulus features were defined and computed, and then how the TRFs were estimated from the data.

The general analysis pipeline, along with stimulus representations and analysis methods are presented in the diagram of figure 1.

2.3.1 Stimulus representation

In light of the current literature, we focused on rule-based syntactic features on the one hand, following cortical tracking of hierarchies suggested by Ding, Melloni, et al. (2016), and on statistical features reflecting predictive processing of sequences as in Weissbart, Kandylaki, and Reichenbach (2020). The syntactic features were computed using the Stanford parser (Manning et al., 2014). The statistical features were computed using the GPT2 language model (Radford et al., 2019).

Importantly, we do not assume that language processing is supported by information-theoretic metrics in the hard sense. That is, we are not proposing a theory of language understanding from mere surface statistics. However, following

the hypothesis given by theories on predictive processing in various cognitive domains and notably for perception, we assume that the brain can extract information-theoretic features from the stimulus, regardless of the specifics of its representational format (e.g., functionally semantic or syntactic), and that these features are actively used to predict internal representations. Those information-theoretic features are thus a proxy for the underlying domain-general cognitive process of predictive inference, notably for predictive processing and Bayesian inference, where surprisal becomes a proxy for prediction error (Peter W Donhauser and Baillet, 2019; Koelsch, Vuust, and Friston, 2019; Federmeier et al., 2007) and entropy for uncertainty (Friston, 2015). Both of these are critical concepts in predictive processing. These quantities are involved in most predictive mechanisms, such as what Andrea E Martin (2016) and Andrea E. Martin (2020) proposed, where cues are extracted at different representational levels to predict or infer upcoming linguistic information.

On the other hand, rule-based features are derived from specific instantiations of parsing mechanisms (such as constituency trees in context-free grammar), which already support a particular theory of syntax derived from linguistics rather than neuroscience. Again, we are not suggesting by the use of those particular features that the brain is precisely implementing such a parsing strategy to compute syntactic representations. Nevertheless, the known sensitivity to such structures (Bai, A. S. Meyer, and Andrea E Martin, 2022; Jonathan R Brennan and Andrea E Martin, 2020; Coopmans, De Hoop, Kaushik, et al., 2022; Coopmans, De Hoop, Hagoort, et al., 2022; Ding, Melloni, et al., 2016; Hale et al., 2022) motivates the use of a metric that captures the general shape of syntactic trees throughout naturalistic sentences. We decided to create a set of features that tracks the complexity of such structures and also captures the integrative mechanisms at play when words or phrases need to be integrated into a larger syntactic unit.

- Syntactic features: To build our syntactic features, we first ran a tree parser on every sentence from our stimuli.
    1. Depth: Syntactic depth is a proxy for syntactic complexity. A word highly embedded within nested structure, that is a word deep in the tree structure hierarchy, will carry a higher value for this feature.
    2. Close: This refers to the number of subtrees being closed at a given word. Some words do not close any subtree, and some will close several at once. This feature encompasses the variability accounting for integrative mechanism such as "merge". When words, or phrases need to be grouped into a larger syntactic unit this feature is incremented. It is also referred to as *bottom-up* count of syntactic structures. As opposed to the top-down count which enumerates opening nodes.
- Statistical features: Extracting the information-theoretic values requires an estimate of the probability distribution of each word in the utterance conditioned on the previous word (over the entire vocabulary). This was quantified using GPT2, a state-of-the-art language model, which is trained specifically to causally predict word from sequences of textual data.
    1. Surprisal: The negative natural logarithm of the probability of a given word item conditioned on the sequence of preceding words $P(w_i|w_{i-1}, w_{i-2}, ..., w_1)$.
    2. Entropy: Mathematically, this is the expected surprisal. It quantifies the amount of uncertainty, at a given word, in predicting the next word.

The correlation is not excessively high and should not, presumably, lead to collinearity issues in the subsequent analysis. To verify we computed the Variance Inflation Factor for each feature. All VIFs were below 2, which is well below the threshold of 5, indicating that there is no collinearity issue in our data (Zuur, Ieno, and Elphick, 2010).

2.3.2 Temporal response functions

With the speech representation time-aligned to the magneto-encephalographic recordings, we then compute optimal filters that map the stimulus to the MEG signals. This is known as forward modelling, or encoding, and in particular the present method boils down to extracting *temporal response functions* of the above features. This method assumes a convolutional linear model mapping from stimulus to MEG data. Such an approach has been successfully applied to recovering brain responses to sound envelope (Keitel et al., 2016b; Etard and Reichenbach, 2019), allowing to further decode auditory attention, and also in estimating brain responses to ongoing linguistic features (Brodbeck, Presacco, and Simon, 2018; Broderick et al., 2018; Weissbart, Kandylaki, and Reichenbach, 2020). This has been applied with different neuroimaging data such as fMRI, ECoG, and M/EEG.

We model the MEG response signal at sensor i (or source estimate location) $\{y_i\}_t$ as a convolution between a kernel β (to be estimated) and the stimulus representation signal $\{x_j\}_t$ for the jth feature. Activity from the data that is not captured by this model is supposed to be Gaussian noise.

$$y_i(t_n) = \sum_{j=1}^{N_{feat}} (\beta_{ij} * x_j)(t_n) = \sum_{j=1}^{N_{feat}} \sum_{\tau=1}^{\tau_{max}} \beta_{ij}(\tau) x_j(t_n - \tau) + \epsilon_{in} \tag{1}$$

$$\hat{y_i}(t_n) = \sum_{j=1}^{N_{feat}} \sum_{\tau=1}^{\tau_{max}} \hat{\beta_{ij}}(\tau) x_j(t_n - \tau) \tag{2}$$

Note that the *hat* symbol, $\hat{}$, correspond to the *estimated* or reconstructed values. It is easy to rewrite equation (2) in a vectorized form as $\hat{\mathbf{Y}} = \hat{\beta}\mathbf{X}$ where the temporal dimension is expressed as column vectors and each channel data (vector) are concatenated along the row dimension of the matrix $\hat{\mathbf{Y}}$. In this formulation, $\mathbf{X}$, called the design matrix in the context of linear models, contains all the lagged time series of every feature in its columns and time samples along its rows: $\mathbf{X} \in \mathbb{R}^{(N_{feat} \cdot \tau_{max}) \times N_{samples}}$. The matrix $\hat{\mathbf{Y}}$ contains the reconstructed MEG data, with the same number of rows as $\mathbf{X}$ and as many columns as MEG sensors or source estimates: $\hat{\mathbf{Y}} \in \mathbb{R}^{N_{samples} \times N_{sensors}}$. Finally, $\hat{\beta}$ is the matrix of estimated TRFs, with a row for each lag for each feature and as many columns as MEG sensors or sources: $\hat{\beta} \in \mathbb{R}^{(N_{feat} \cdot \tau_{max}) \times N_{sensors}}$.

Written in its vectorized form, we can easily see how the closed-form formula to estimate $\hat{\beta}$ arises:

$$\mathbf{Y} = \mathbf{X}\beta \tag{3}$$

$$\mathbf{X^T Y} = \mathbf{X^T X}\beta \tag{4}$$

$$\hat{\beta} = (\mathbf{X^T X})^{-1}\mathbf{X^T Y} \tag{5}$$

$$\tag{6}$$

However, when the inversion of $\mathbf{X^T X}$ is unstable, which easily happens with continuous predictors, a regularised autocorrelation matrix is used instead (via Thikonov regularisation):

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} \tag{7}$$

where λ is the regularisation parameter, and $\mathbf{I}$ is the identity matrix. Wherever used, the regularisation parameter was set to $\lambda =< \lambda_k >_k$, the arithmetic mean over the eigenvalues of $\mathbf{X^T X}$.

Inter-word phase coherence and Phase-amplitude coupling A first analysis for estimating phase amplitude coupling consisted in scanning across several frequencies using a time-frequency representation of the MEG data. We used the Morlet wavelet transform to compute the time-frequency decomposition of the MEG data. The Morlet wavelet transform was computed using seven cycles between 1 and 80Hz. We extracted the average power by taking the absolute of the complex Fourier coefficients and the phase by normalising the complex wavelet output by its absolute value.

One challenging aspect in computing the inter-trial phase consistency (ITPC, or inter-event phase clustering and in the case of the present study: inter-*word* phase clustering) is that it relies on the clustering of phase, evidently, across trials (defined by word onset timings). Hence, it depends on trial-based experimental design. We propose a novel approach to circumvent trial design and leverage the use of naturalistic stimuli by adapting the forward linear model to compute an equivalent of ITPC and PAC for continuous M/EEG recordings.

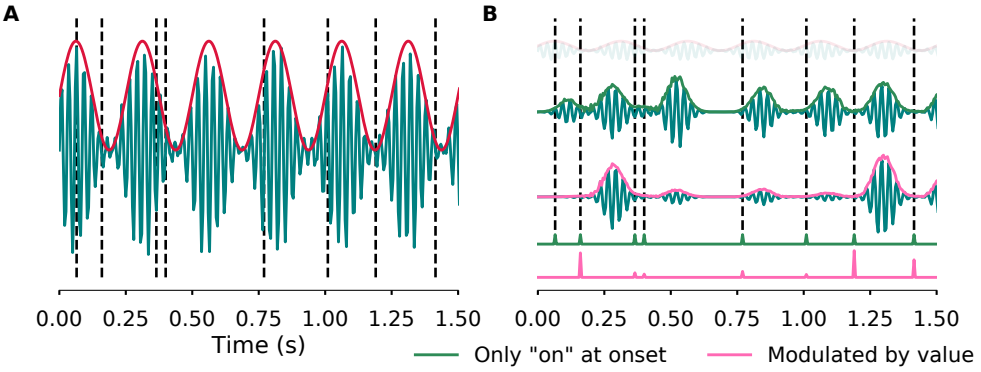
Our approach consists in using the complex analytical signal to compute a complex-valued TRF which will jointly contain phase and amplitude information. Computing the TRF is equivalent to finding the kernel of the convolution, which triggers the recorded response and thus captures and summarises the phase concentration from a continuous recording. In essence, the TRF computed from a comb-like time-series as features is analogous to ERP analysis. We can extend this analogy and consider this method as a way to compute ITPC and PAC from continuous recordings.

The classic, trial-based, ITPC is computed as:

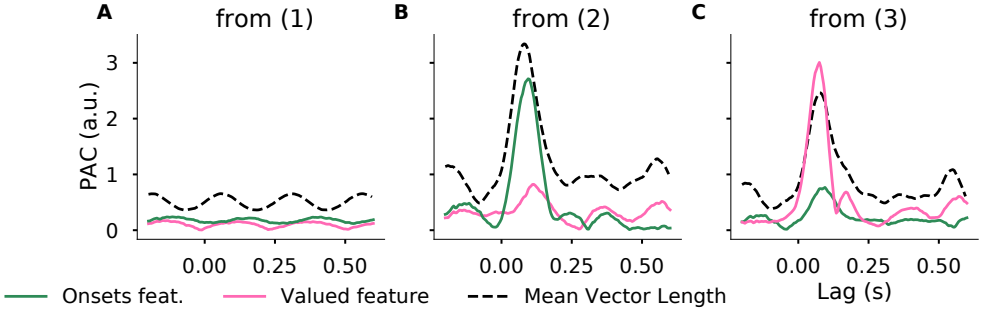
$$ITPC_i(t) = \frac{1}{N_{trials}} \left| \sum_{trials} e^{\phi_{i;\delta}(t)} \right|$$

While the PAC can be computed as (Canolty et al., 2006):

$$PAC_{\delta \to \beta}(t) = \frac{1}{N_{trials}} \left| \sum_{trials} r_\beta(t) e^{\phi_{i;\delta}(t)} \right| \tag{8}$$



(a) Simulated phase-amplitude coupling: The left panel shows a simple phase-amplitude coupling where the slow phase of the signal in red modulates the amplitude of a faster signal in green. In the right panel, the amplitude is modulated differently depending on the timing and values of an external feature signal. From top to bottom: same as in A (phase is not tight to event timing), phase is locked to event timing, phase is locked to event timing and feature value modulates the strength of the coupling.



(b) Estimation of simulated phase-amplitude coupling using TRF (in colour: pink and green) or with an epoch-based approach, i.e. computing the mean-vector length across trials (dashed black curves). Each panel from left to right corresponds to the three kind of simulation as in B above.

Figure 2: PAC simulations. We can recover the effect of an individual feature affecting phase-amplitude coupling from the linear forward model.

Therefore, we can use the continuous signals $e^{\phi_{i;\delta}(t)}$ and $r_\beta(t)e^{\phi_{i;\delta}(t)}$ to compute ITPC and PAC respectively, with respect to different features. Using only word onset (a feature that is one at word onset and zero otherwise) is equivalent to the trial-based computation of those quantities. However, this extended framework allows us to incorporate other exogenous variables and thus analyse their respective contributions to ITPC and PAC. Forte, Etard, and Reichenbach (2017) have successfully applied such a modelling approach on complex signals to recover phase and amplitude response in brainstem recordings.

In order to investigate whether this method reliably recovers features that specifically contribute to some phase-amplitude coupling, we simulated the effect and recovered the dynamics of coupling from features, beyond the event-related mean vector length. Those simulations are presented in figure 2. We simulated three different scenarios: (A) a simple phase-amplitude coupling where the phase of low-frequency signal modulates the amplitude of a faster signal regardless to event timing. Then we simulated phase-amplitude coupling occurring in a time-locked manner after some events. Finally, a third simulation was performed where phase-amplitude coupling occurs in a time-locked manner and the amplitude is modulated differently depending on the values of an external feature signal. We then computed the TRF for each of those simulations and compared them to the mean vector length computed across trials. As seen in figure 2, we can recover the effect of an individual feature affecting phase-amplitude coupling from the linear forward model.

2.3.3 Statistics

Power spectral densities and coherence comparison were carried out using paired t-tests, correcting for multiple comparison by controlling for the false discovery rate via the Benjamini-Hochberg procedure implemented by MNE Python `fdr_correction` method. Reconstruction scores and kernel coefficients learned through linear regression were tested for significance against a null model of temporal response functions. We constructed an empirical estimate

of the null distribution represented by a model where syntactic and predictive features were unrelated to the stimulus. This is realised by shifting the values of each feature while keeping the word-onset timings intact. Therefore, each reconstruction score can be compared to the corresponding null model which matches the number of features. The relative score difference for each subject is then evaluated using one sampled t-test, comparing their value against zero. For TRF time-courses, we used a cluster permutation test to assess significance conditions (actual versus null models) and extract spatio-temporal clusters (Oostenveld et al., 2011).

3 Results

We measured power spectral density (PSD) of the MEG data in order to assess the quality of the data and the presence of the expected neural oscillations. As seen in figure 3, panels A and B, the MEG data presents neural oscillations in the alpha and beta bands. We found marginally significant difference in power between the French and Dutch listening conditions in the beta band (FDR corrected for multiple comparison across all frequencies, corrected p-value = 0.06, one sampled t-test, dof = 24). We then computed the cerebro-acoustic coherence by taking the magnitude coherence squared between the sensor-level MEG signals and the sound envelope of the stimuli, see figure 3 panel C. This reveals coherent phase alignment between sound amplitude and MEG signals in the delta and theta bands. Importantly, while those reflect processing of speech-like sounds they do not directly reflect comprehension as it occurs for both French and Dutch listening conditions. Besides, it is actually significantly greater for the French condition in the delta and theta-band (p = 0.01, cluster-based permutation test using one sample t-test as statistic). Finally, we computed a time-frequency representation for every word-epochs. We then extracted power modulation and inter-word phase clustering from the complex Fourier coefficients. As seen in figure 3 panel D, we observe power modulation in the beta band, and inter-word phase clustering in the delta and theta band.

3.1 Joint contributions of syntax and statistical features

We first analysed the spatio-temporal dynamics of low-frequency activity in response to all features using linear forward models. We assessed the reliability of those features in representing the MEG data by computing the reconstruction accuracy at each sensor (or sources) for a model containing the feature of interest. As seen in figure 4, each set of features generates a significant increase in reconstruction accuracy compared to null models. Importantly the full model, which contains both statistical and syntactic features, shows a significantly higher reconstruction accuracy than both other models.

3.2 Time-resolved contributions of syntax and statistical features

We further investigated the time-resolved contribution of each feature set to the overall score. We computed the score for each feature set for a series of time windows, and computed the relative contribution of each feature set to the overall score. This is shown in figure 4 for the delta band and beta band. Most of the variance is explained solely by the time-locked activity with respect to word onsets. However, results show that syntactic features mostly contribute to the score throughout a large time window extending at earlier and later time lags. For both feature, we see significant reconstruction beyond and above word-onset only models around 0ms. We attribute the improvement around word onset to predictive mechanisms for both syntactic and statistical features. Finally, the improvement in reconstruction accuracy for syntactic features at later time lags is likely due to the integration of words into larger syntactic units. This is supported by the fact that TRF time course for closing bracket count feature present the largest coefficients at later time-lags. This feature corresponds to the number of syntactic units that are closed at a given word, which is a proxy for the integration of words into larger syntactic units.

3.3 Phase consistency and power modulation at word onsets

Evidence from recent studies motivate the analysis of phase-locked activity in the delta and theta band (Peelle and Davis, 2012; Keitel et al., 2016b; Etard and Reichenbach, 2019; Jonathan R Brennan and Andrea E Martin, 2020). Altogether, we hypothesised change in phase consistency in delta and theta range after word onset, potentially coupled with modulation of beta oscillatory activity and with gamma broadband activity (Luc H. Arnal, Keith B. Doelling, and Poeppel, 2015; Pittman-Polletta and Kocsis, 2022). To further validate this choice of frequency bands of interest, we computed word-triggered phase consistency and power modulation. Figure 3 presents the long-term power spectral density across all stories (A) as well as power modulation and inter-trial phase consistency happening after word onsets (D). We could retrieve those word-related modulation by computing a TRF model with word onset only and band passed data (see figure 4). We found significant power modulation in the beta bands and significant inter-word phase consistency in the delta and theta bands (figure 4, panel C).

3.4 Phase amplitude coupling

First, we tested the presence of PAC across the entire signal, regardless of the timing of words. That is, we used equation (8) but instead of averaging across trials, the average over time is used. We found evidence of PAC between

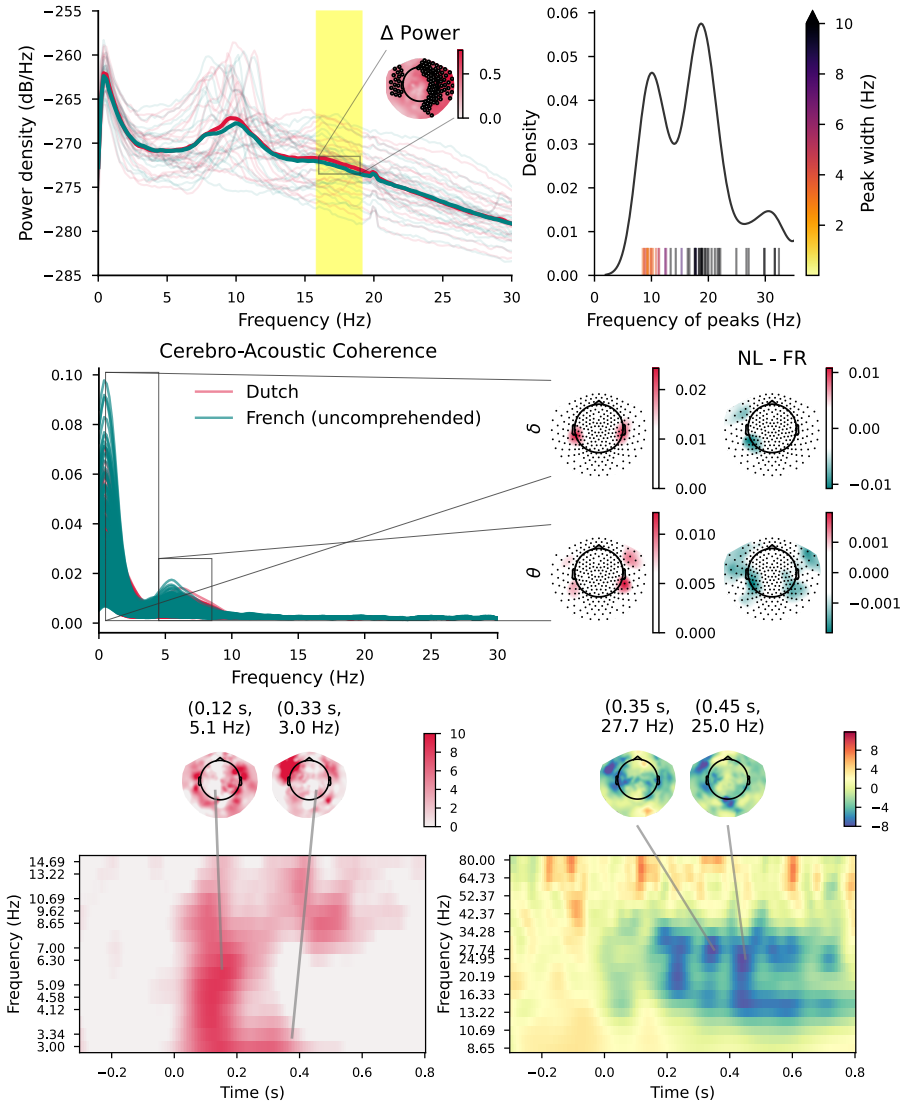


Figure 3: Power spectral density of the MEG data. A and B: The left panel shows the PSD for the French and Dutch listening conditions, the right panel shows the distribution of peak frequencies colour-coded by the width of the peak for all subjects. Peaks are detected on the average sensor-level PSDs. C: cerebro-acoustic coherence. We computed the magnitude squared coherence between MEG sensor data and sound envelopes. Topographic plot on the right indicate average coherence values in the given regions as well as the difference contrast between Dutch and French conditions. E and G: Power modulation at word onset and ITPC respectively.

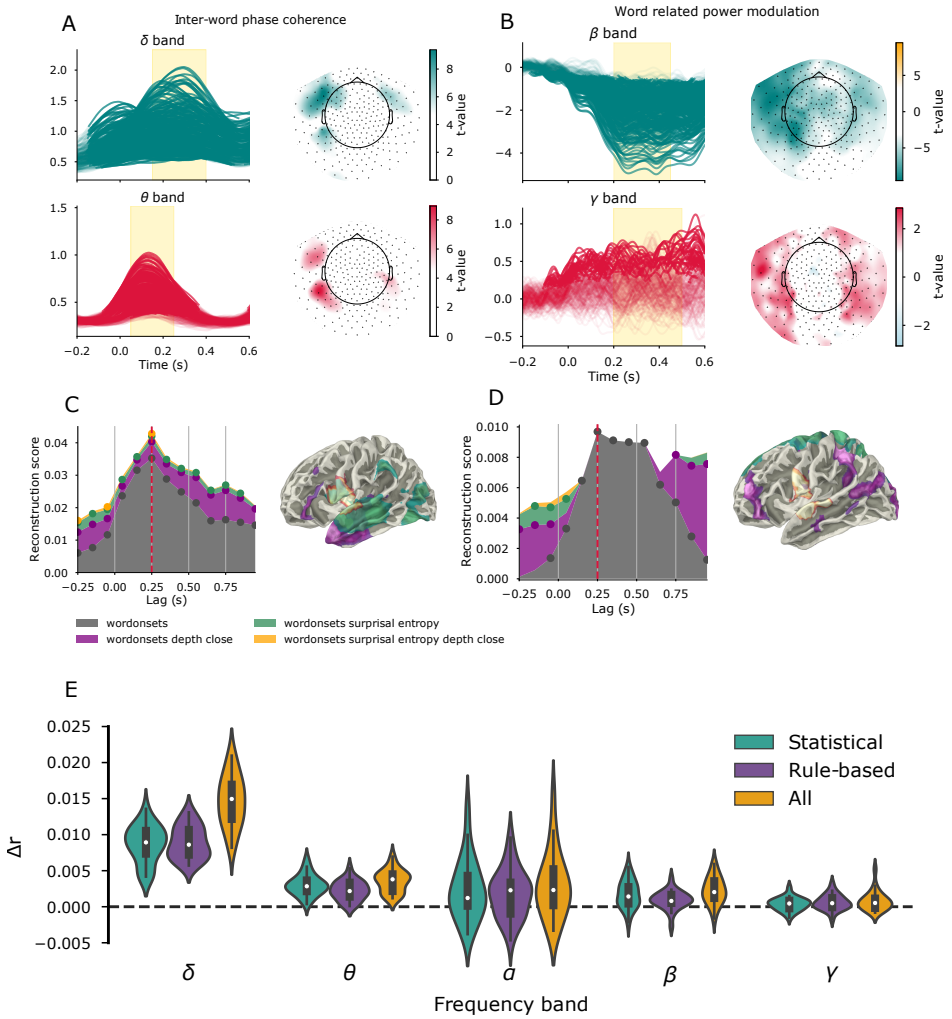


Figure 4: A: ITPC computed using TRF model with word onset only on band passed complex data, in bold are the spatio-temporal clusters greater than zero, the shading yellow area indicate the time segment over which the average is taken to display the topographic plot. B: Power modulation computed on band-passed power using a TRF model with word onsets only. C & D: Reconstruction accuracy for different time windows, with relative contribution of each feature set displayed as stacked area plot. On the right sub-panel is displayed the spatial distribution of the score for each feature set. C is for the delta band, and D for the beta band. E: Scores increment (over null models scores) for each feature sets and frequency bands. For the delta and theta band, we computed the ITPC, while for higher frequency bands, the power were taken.

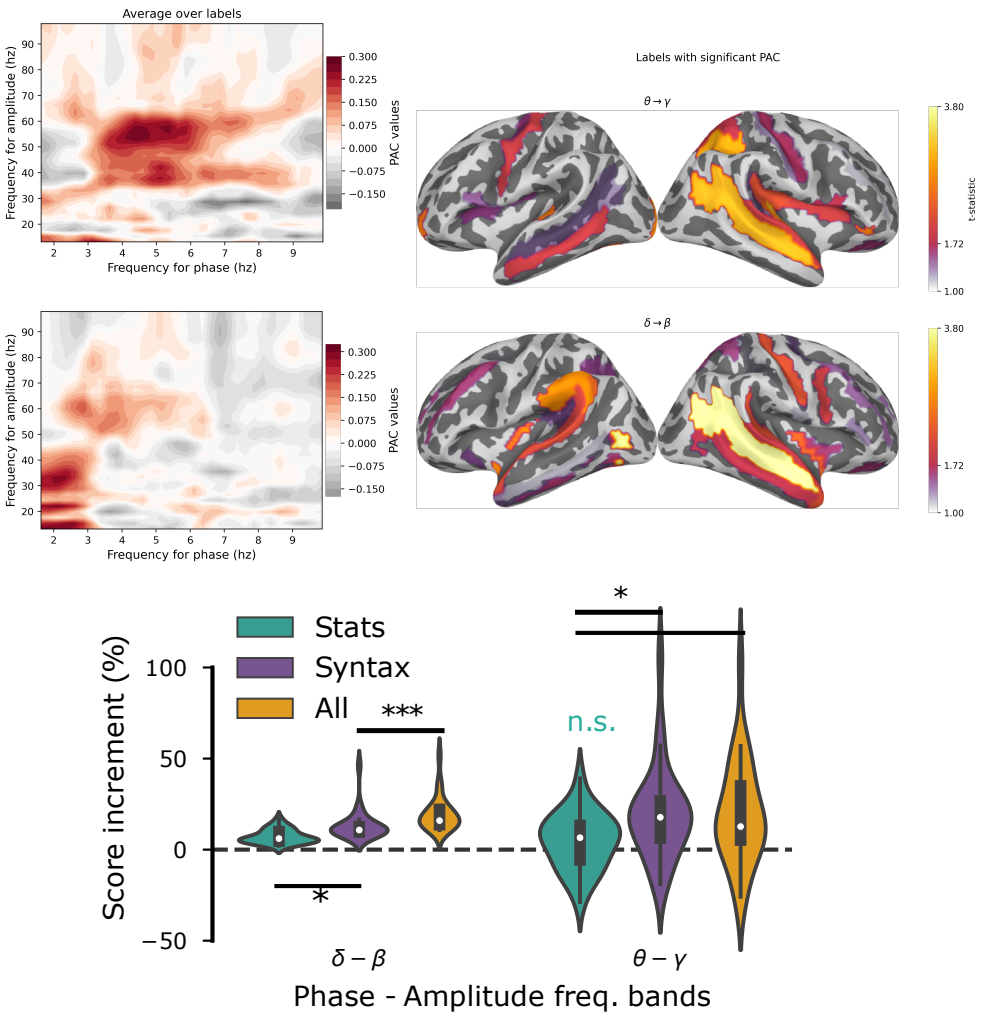


Figure 5: The top panels present the stimulus-wide PAC, not time-locked to word events. PAC values were normalised (z-scored) against PAC estimated from surrogate data (shuffling amplitude segments). As every frequency across phase and amplitude are scanned, PAC was only computed for labels of a parcellation of the source model to reduce computation time and memory resources. The bottom panel shows the scores of the TRF-based PAC models for different feature sets. Paired t-tests were used to compare the scores of the different models (dof = 24, stars indicate a quantisation of the exponent in power of 10 of the p-value).

the delta phase and the power in the beta and gamma range, as well as between the phase of theta range activity and gamma power (figure 5). PAC estimates were compared to surrogate data to assess significance. We then introduced a new method to disentangle the effect of different stimulus features to PAC. As such, we are investigating PAC after word onset. The simulations described in figure 2 show that we can recover the effect of an individual feature affecting phase-amplitude coupling from a linear forward model.

The new method relies on a linear forward model akin to temporal response function computation where instead of the band-passed MEG data we used the normalised analytical signal. By doing so, we effectively look at how the phase of the complex analytical signal clusters at a given lag. The advantage of this method is that it takes into account every predictor we feed into the forward model and thus allows us to establish which features weigh in the most in the clustering of phase. Using surrogate data based on an arbitrary phase resetting mechanism, we could show that we indeed recover value comparable to the classic inter-trial phase clustering method, but moreover that we could distinguish which feature was responsible for the phase reset. We found significant reconstruction accuracy for those models (see figure 5, panel B).

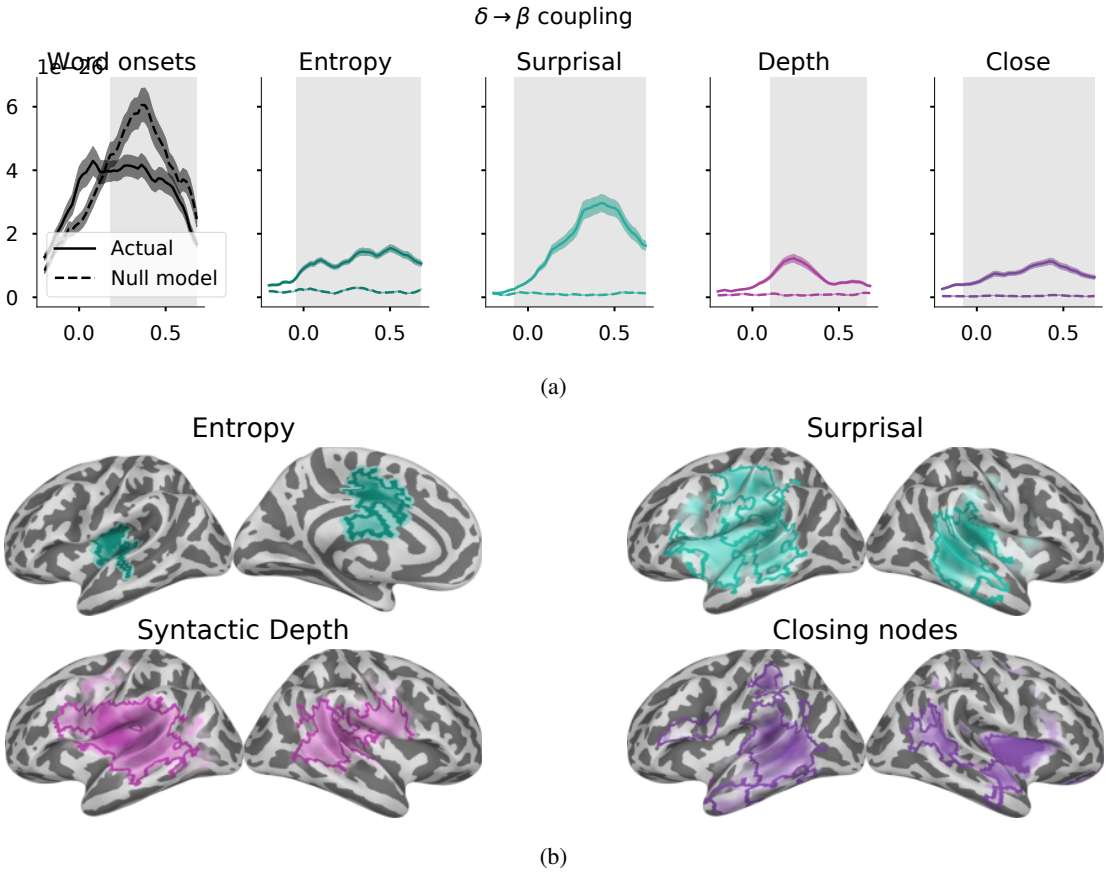


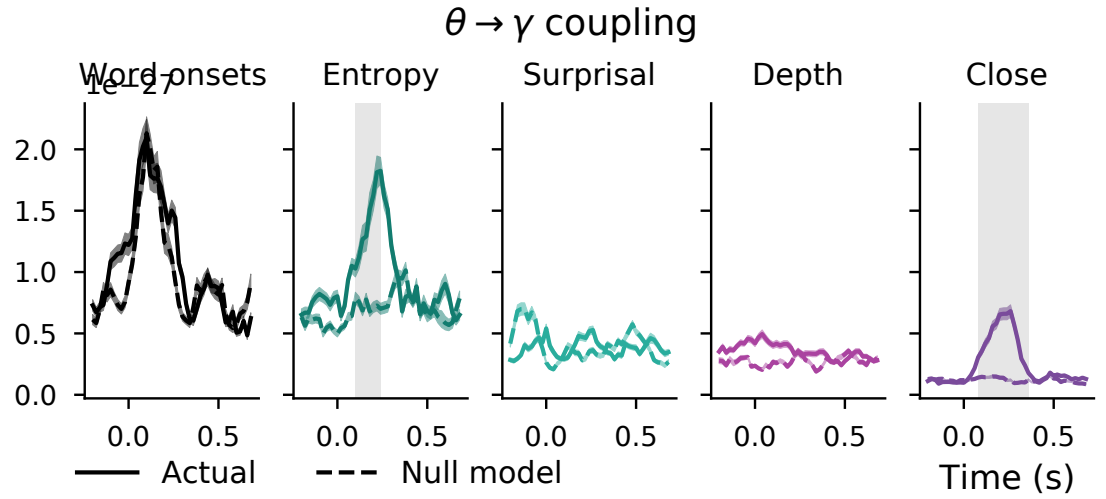
Figure 6: PAC between delta and beta activity. (a): Time course and topographies. (b): Significant clusters in source space

Fig. 6 and 7 shows results of phase-amplitude couplings. All features significantly modulated cross-frequency coupling in the delta to beta bands. While only precision entropy and closing bracket counts showed a significant modulation of gamma power through theta phase. Higher PAC coefficients were found in the superior temporal gyri in bilaterally. In the left hemisphere, we found significant clusters in the inferior frontal gyrus, the anterior temporal lobe and the superior temporal gyrus. In the right hemisphere, we found significant clusters in the inferior frontal gyrus, the anterior temporal lobe and the superior temporal gyrus. The left inferior frontal gyrus and the left anterior temporal lobe are two regions that have been previously associated with syntactic processing and semantic composition.

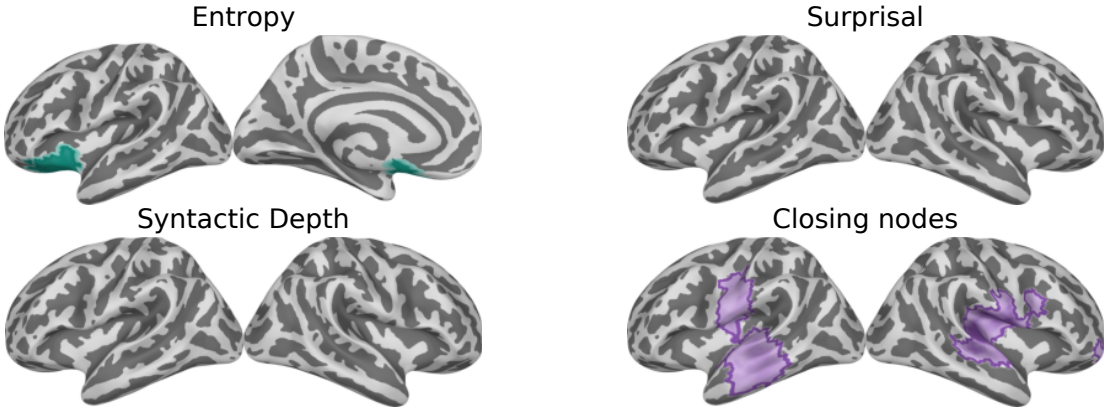
4 Discussion

A growing number of studies demonstrated how to measure the sensitivity of the brain to naturalistic speech for word-level features. It remains difficult to control for confounding aspects, such as in Ding, Melloni, et al. (2016), where it has been argued that the chunking observed around syntactic phrases might be elicited simply by word-level occurrence statistics or by the repetition of part-of-speech tags (Frank and Yang, 2018; Ten Oever, Kaushik, and Andrea E Martin, 2022).

We built a word-level representation of naturalistic speech encompassing both syntactic, rule-based, features and statistical, data-driven, features. We showed that both feature sets could be recovered from MEG signals and that they could be used to decode the comprehension state of the subject. Finally, we showed how phase-amplitude modulation jointly occurs for both feature sets, suggesting that they are both processed and orchestrated in parallel. We found overlapping brain regions for both feature sets, in particular in the left inferior frontal gyrus, and in the anterior temporal lobe, two brain regions that have been previously associated with syntactic processing and semantic composition (Bemis and Pylkkänen, 2011; Andrea E. Martin and McElree, 2011; Ding, Melloni, et al., 2016; Pylkkänen and Jonathan R. Brennan, 2019; Pylkkänen, 2020; Nelson et al., 2017; Zaccarella, Schell, and Friederici, 2017).



(a)



(b)

Figure 7: PAC between theta and gamma activity. (a): Time course and topographies. (b): Significant clusters in source space

Using naturalistic stimuli, we dissociated, on the one hand feature set computed solely based on statistics of word sequences as computed via GPT model and, on the other hand a set of rule-based abstract features built directly from constituency tree structures. Temporal response functions obtained from those feature sets could each explain variance in MEG signals beyond chance and gave enhanced representations of the signal that allowed for decoding of comprehension state on unseen subject data.

4.1 Rule-based and predictions-based syntactic features are processed in parallel

Nelson et al. (2017) showed that broad-band gamma activity was more sensitive to syntactic surprisal, dissociating statistical features from syntax. Although, they trained statistical models purely on sequences of part-of-speech tags, which conveniently remove the lexico-semantic properties of the context. In the present study, we cannot control for the syntactic information contained in the statistical features, but we verified that the sensitivity to phrasal boundaries is at least minimal in those features. In particular, the language model is trained to predict the next word given the previous ones, and thus may learn an approximation of syntactic distributions across words as much as that approximation facilitates next word prediction. Therefore, the statistical features are not fully independent of syntax (Slaats and Andrea E Martin, 2023). However, syntactic features are computed from annotations of the constituency-parse tree structure of the sentences, and thus do not contain any information about the statistics of word sequences. And we assume that they are solely driven by the syntactic structure of the sentences.

The result of the joint model explaining over and beyond any of the individual feature sets suggests that the brain is sensitive to both aspects of language in an inseparable manner. This synergy is observed in regions where both types of information are encoded with equal strength. However, we observed in several frequency bands a dissociation between the two feature sets. In particular, we found that the statistical features were more sensitive to the theta band, while the syntactic features were more sensitive to the delta band. This dissociation is in line with previous studies showing that the theta band is related to the processing of statistical regularities (Giraud and Luc H Arnal, 2018; Peter W. Donhauser and Baillet, 2020), while the delta band is related to the processing of syntactic structure (Bai, A. S. Meyer, and Andrea E Martin, 2022; Bastiaansen, Magyari, and Hagoort, 2010; Jonathan R Brennan and Andrea E Martin, 2020; Coopmans, De Hoop, Hagoort, et al., 2022; Ding, Melloni, et al., 2016; Kaufeld, Bosker, et al., 2020; L. Meyer and Gumbert, 2018b; L. Meyer, 2018; L. Meyer and Gumbert, 2018a; L. Meyer, Sun, and Andrea E Martin, 2020b; L. Meyer, Sun, and Andrea E Martin, 2020a; Nelson et al., 2017; Slaats, Weissbart, et al., 2023; Ten Oever, Carta, et al., 2022; Zioga et al., 2023). This pattern where statistical information aids the inference of categorical information can be observed at other levels of granularity during speech processing, for instance, in the processing of the spectro-temporal acoustic features into categorical phonemes, and from phonemes themselves into lexical representations - in this sense, integration from phonemes to words, is modulated by statistical cues from top-down predictions (Giraud and Luc H Arnal, 2018; Peter W. Donhauser and Baillet, 2020; Tezcan, Weissbart, and Andrea E Martin, 2023). In particular, our findings relate to Giraud and Luc H Arnal (2018) who showed that slow cortical activity is related to the processing of the predictability of the next word in a sentence. Our results extend this finding by showing that low-frequency phase information modulates gamma power in a way that is related to the predictability of the next word in a sentence, namely when the uncertainty of those predictions is high.

4.2 Brain correlates of syntactic computation

Different aspects of syntax have been previously linked to brain activity ranging from syntactic violation and the P600 event-related potential component (Molinaro, Barber, and Carreiras (2011), for a review, see Kuperberg (2007)) to phrase structure tagging (Ding, Melloni, et al., 2016) and syntactic surprisal (Nelson et al., 2017). In particular, the *merge* operation (Bastiaansen, Magyari, and Hagoort, 2010; J. Brennan et al., 2012; Coopmans, Kaushik, and Andrea E Martin, 2023; Nelson et al., 2017) has the privilege of reflecting a crucial aspect of syntax, namely its recursive and combinatorial power. The merge operation is fundamental for composition and syntactic unification. It is also crucial in endowing language with its recursivity, allowing it to generate an unbounded set of utterances from a finite vocabulary set. Researchers have focused on pinpointing a specific location for such computation (e.g., Zaccarella, Schell, and Friederici, 2017). However, it is unclear whether such computation, in the brain and in complementary fashion to computational-level claims in formal theories, occurs in isolation from semantic processing and more generally, from the dynamical update and integration of cues at different levels of the speech and language processing hierarchy (Andrea E Martin, 2016; Andrea E. Martin, 2020; Ten Oever, Carta, et al., 2022). We used the closing bracket count as a proxy for the binding operation, similar to Nelson et al. (2017) and Andrea E Martin and Doumas (2017). We found that spatio-temporal responses to both syntactic features are generally left-lateralised, but not exclusively so (see also Bai, A. S. Meyer, and Andrea E Martin, 2022); moreover, we found distributed activity in the left anterior temporal lobe and in parietal regions and inferior frontal gyrus. In the time domain, we found a significant effect of closing nodes in both delta and theta bands. Other studies have attributed neural tracking to speech in the theta-band to the active chunking process of word units (Teng et al., 2018), perhaps aided by predictions (Giraud and Luc H Arnal, 2018; Peter W. Donhauser and Baillet, 2020). In that respect, our data suggest that syntactic processing may also play a role in aligning with linguistic units and in providing a basis for tracking of such linguistic units (see also Ten Oever and Andrea E Martin, 2021).

4.3 Phase-amplitude coupling for binding latent representations

Syntactic and predictive processing do not need to be fully disjoint in the brain. Each feature can jointly contribute to the prediction of upcoming input while at the same time help in integrating context with currently processed units. Locally, in a given neural circuit, we propose that binding of neural representations can be made through phase-amplitude coupling (see Canolty et al. (2006), Doumas, Hummel, and Sandhofer (2008), Andrea E Martin and Doumas (2017), Andrea E Martin and Doumas (2019), Andrea E Martin and Doumas (2020), and Andrea E. Martin (2020) for a review). This mechanism has been proposed to be a general mechanism for neural binding (Canolty et al., 2006; Von der Malsburg, 1995; Von der Malsburg, 1999), and has been argued to be involved in the processing of linguistic features (Bai, A. S. Meyer, and Andrea E Martin, 2022; Giraud and Luc H Arnal, 2018; Peter W. Donhauser and Baillet, 2020; Andrea E. Martin, 2020). In particular, we found that the coupling between theta phase and gamma amplitude was stronger, specifically for entropy and bottom-up integration of nested trees. This suggests that the two feature sets are processed in parallel. These features reflect how neural dynamics anticipate precision-weighted predictions and integrate a word to the current phrase.

We see evidence for an oscillatory mechanism from the simple word-evoked activity in both phase constituency and induced power across all canonical frequency band of M/EEG. Although we did not find significant gamma power

modulation and note, importantly, that the phase constituency may reflect mere evoked and time-locked activity. Interestingly we see a typical decrease in beta power at the word-level. But once we disentangle the effect of specific features, we find for instance, a positive effect of close, or phrasal integration, which is in line with the structure and sentence-building effect leading to increased beta power observed in Bastiaansen, Magyari, and Hagoort (2010), Ashley Glen Lewis, Schoffelen, et al. (2017), and Ashley Glen Lewis, Wang, and Bastiaansen (2015). More recently, an analysis on the dependency parse of sentences on this very same dataset also provided with evidence for the role of beta for maintenance and/or prediction during dependency resolution. Zioga et al. (2023) showed that the beta power was modulated proportionally with the number of dependencies to be resolved. This would result in a consistent power increase while complex sentences are being processed. We also find a positive effect of depth in the delta band, which is in line with the syntactic processing account of delta tracking (Bai, A. S. Meyer, and Andrea E Martin, 2022; Bastiaansen, Magyari, and Hagoort, 2010; Jonathan R Brennan and Andrea E Martin, 2020; Coopmans, De Hoop, Hagoort, et al., 2022; Ding, Melloni, et al., 2016; Kaufeld, Bosker, et al., 2020; Nelson et al., 2017; Ten Oever, Carta, et al., 2022; Slaats, Weissbart, et al., 2023; Zioga et al., 2023).

Going further, we suggest that those results supports the dual timescale for predictive speech processing as proposed by Peter W. Donhauser and Baillet (2020). Indeed, the observed theta/gamma coupling occurs for highly expected information gain (high entropy leading to stronger coupling) and for words which must integrate into a larger number of constituents. This is in line with the idea that theta is a sensory sampling mechanism tuned to maximize expected information gain (Peter W. Donhauser and Baillet, 2020). A word that closes several nested constituents will generally be well predicted syntactically but highly anticipated semantically, as it plays a crucial role in the sentence (and even more possibly in word-final language). As such, an externally driven, weakly entrained, oscillator could synchronize the integration of a word into a larger syntactic structure while aligning through phase synchrony and nested theta-gamma oscillations the excitability of neural assemblies engaged. This is synchronized both with the input and the internally generated temporal predictions further as the expected information gain is high. The computational model proposed by Ten Oever and Andrea E Martin (2021) also supports such framework where the predictability of a word will generate a corresponding phase advance or lag. Although in this our study, it is precisely the expected information gain that is driving the phase synchrony. On the other hand, delta/beta coupling occurs for words that are less predictable, i.e., with a high surprisal value. This is in line with the idea that delta is encoding non-redundant information (Peter W. Donhauser and Baillet, 2020). That is, novel information, deviating from internally generated predictions must *update* the internal model. Top-down predictions, but also updates, have been already linked to beta power modulation. All in all, this mechanism, if supported via delta-phase, is bound to endogenously generated predictions and thus to the internal model of the listener (Andrea E. Martin, 2020). Given the synchronous increase in coupling for syntactic features (depth and close) we suggest that this internal model update is aligned with the slower rhythm of (predicted) phrasal boundaries (L. Meyer, Sun, and Andrea E Martin, 2020b).

4.4 Towards a predictive coding theory for language processing?

If based purely on data showing tracking to sound envelope, we must note that delta band does not reflect purely linguistic processes. First we see on the present data that coherence with sound envelope is stronger in delta for uncomprehended language, and that power alone does not carry sensible difference between our listening condition (figure 3, panel A and B). Moreover, low-frequency tracking of envelope is sensitive to acoustic changes too, rather than linguistic information, for instance it becomes larger when pauses are inserted in the stimulus (Deoisres et al., 2023). Here however, we note the role of delta phase and its coupling with higher frequency band, in particular beta, for the processing of linguistic features. We observed strong dependency of beta modulation, coupled through the phase of delta due to word surprisal. This is in line with the idea that beta carries out top-down predictions and internal model updates while being modulated by the phase of endogenously generated delta rhythms (Ashley Glen Lewis, Wang, and Bastiaansen, 2015; Spitzer and Haegens, 2017; Zioga et al., 2023). Given the importance of theta- and delta-band speech tracking for comprehension (e.g., Etard and Reichenbach, 2019; Weissbart, Kandylaki, and Reichenbach, 2020; Ding, Pan, et al., 2018; Peelle and Davis, 2012), we propose that the coupling between delta and higher frequency bands is a general mechanism for the processing of linguistic features, possibly leveraging the temporal regularities of lower-level cues to align with internally generated predictions and representations. Previous studies have highlighted the role of beta oscillations for predictive coding (Ashley Glen Lewis, Wang, and Bastiaansen, 2015; Ashley G. Lewis et al., 2016).

We use surprisal as a proxy for information prediction, which may or may not take some syntactic and semantic features into account. It certainly does not build upon lower levels such as spectro-temporal acoustic information, as this would imply having trained a statistical language model from raw audio data. In the classical view of predictive processing though, bottom up prediction errors are matched against top-down predictions. One caveat in language is that, beyond the sensory representations, the hierarchy of representation is not well-defined. Structural and semantic information may not need to always sit within a strict hierarchy (except for example when function application or domain specification for semantic functions is specified by syntactic structure - which may be quite often), and the

brain may nonetheless leverage both sources of information to work in parallel. Thus, it is unclear what becomes top-down or bottom-up information at the sentence level, when considering only word-level features. Moreover, we must account for the temporality of speech, where contextual information, from previous words, also helps to refine predictions. In this study, surprisal and precision (entropy) encompass such information from the word sequence, while syntactic features are more likely stemming from higher levels sources (i.e., not purely lexically-driven) if we consider the possibility that sentence level representations serve as top-down predictions for upcoming word-level processing. Indeed, the information added by incoming words within nested syntactic structures comes about by taking in account the embedding into a larger constituent, and thus reflects linguistic knowledge and processing over the entire sentence, if not discourse. As such, on a predictive account, the information from lower levels, such as phoneme sequence and fine-grained acoustic information, together with their relation to word recognition, must also be accounted for to fully capture the predictive context of speech processing. A comprehensive account of a predictive coding theory of speech processing must therefore incorporate an ongoing model of language processing together with perceptual processing of incoming sensory input.

5 Conclusion

Syntactic rules, observed in the abstract formalism of constituency trees and context-free grammars, walk hand in hand with surface statistical features to construct meaningful representations, and lead from speech processing to language comprehension in the brain.

We investigated the contribution of rule-based syntactic features together with the sequential predictability of words to cortical signals. Each set of features resulted together in a more accurate representation of MEG signals, suggesting an overlap in how the brain encodes and processes those different features. By computing a forward model, we could extract the neural response to those features from naturalistic listening conditions, thus leveraging the need to manipulate the stimulus to exhibit responses to phrase structures. Syntactic operations, such as *merge*, operationalised here with the closing bracket count, together with the depth in constituency tree structures, showed a temporally broader response around word onset as compared to the sequence-derived features of surprisal and entropy. Across several frequency bands we observed distinct networks with some overlapping regions were both feature sets improved reconstruction accuracies. In particular, we found that the statistical features were more sensitive to the theta band, while the syntactic features were more sensitive to the delta band.

Phase-amplitude coupling analysis revealed that the two feature sets are processed in parallel, and are used to predict the next segment (lexicalised items in the current study) as it is integrated with the current context. We found that the coupling between theta phase and gamma amplitude was stronger, specifically for entropy and for the integration of nested trees. This suggests that the two feature sets are jointly used to update predictions and integrate the linguistic content to the current phrase. Given the role of theta band in speech tracking (Peter W. Donhauser and Baillet, 2020; Etard and Reichenbach, 2019; Teng et al., 2018) we suggest that such rhythms is tightly bound to the acoustic signal and reflect thus temporal predictions aligned with high expected information (Ten Oever and Andrea E Martin, 2021). More generally, all features presented significant cross-frequency coupling between delta and higher frequency bands. But we observed a larger delta-beta coupling for surprisal, potentially in line with the idea that beta carries out top-down predictions and internal model updates while being modulated by the phase of endogenously generated delta rhythms. Given the importance of theta- and delta-band speech tracking for comprehension (Etard and Reichenbach, 2019; Weissbart, Kandylaki, and Reichenbach, 2020; Ding, Pan, et al., 2018; Peelle and Davis, 2012), we propose that the coupling between delta and higher frequency bands is a cortical computational mechanism for the processing of linguistic features, as speech becomes language, and for leveraging the temporal regularities of lower-level cues to align with internally generated predictions and representations.

Acknowledgments

AEM was supported by an Independent Max Planck Research Group and a Lise Meitner Research Group “Language and Computation in Neural Systems”, by NWO Vidi grant 016.Vidi.188.029 to AEM, and by Big Question 5 (to Prof. dr. Roshan Cools & Dr. Andrea E. Martin) of the Language in Interaction Consortium funded by NWO Gravitation Grant 024.001.006 to Prof. dr. Peter Hagoort. Hugo Weissbart was supported by NWO Vidi grant 016.Vidi.188.029 to AEM.

References

Antonello, Richard and Alexander Huth (2022). "Predictive coding or just feature discovery? an alternative account of why language models fit brain data". In: *Neurobiology of Language*, pp. 1–16.

Arnal, Luc H., Keith B. Doelling, and David Poeppel (Sept. 2015). "Delta-Beta Coupled Oscillations Underlie Temporal Prediction Accuracy". In: *Cerebral Cortex (New York, NY)* 25 (9), p. 3077. ISSN: 14602199. DOI: 10.1093/CERCOR/BHU103. URL: /pmc/articles/PMC4537446/%20/pmc/articles/PMC4537446/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4537446/.

Arnal, Luc H. and Anne-Lise Giraud (July 2012). "Cortical oscillations and sensory predictions". In: *Trends in Cognitive Sciences* 16.7, pp. 390–398. ISSN: 13646613. DOI: 10.1016/j.tics.2012.05.003.

Bai, Fan, Antje S Meyer, and Andrea E Martin (2022). "Neural dynamics differentially encode phrases and sentences during spoken language comprehension". In: *PLoS Biology* 20.7, e3001713.

Bastiaansen, Marcel and Peter Hagoort (Jan. 2006). "Oscillatory neuronal dynamics during language comprehension". In: *Progress in Brain Research* 159, pp. 179–196. ISSN: 0079-6123. DOI: 10.1016/S0079-6123(06)59012-0. URL: https://www.sciencedirect.com/science/article/pii/S0079612306590120.

Bastiaansen, Marcel, Lilla Magyari, and Peter Hagoort (July 2010). "Syntactic Unification Operations Are Reflected in Oscillatory Dynamics during On-line Sentence Comprehension". en. In: *Journal of Cognitive Neuroscience* 22.7, pp. 1333–1347. ISSN: 0898-929X. DOI: 10.1162/jocn.2009.21283. URL: http://www.mitpressjournals.org/doi/full/10.1162/jocn.2009.21283%7B%5C#%7D.VZOWS%7B%5C_%7DlVhBc%20http://www.mitpressjournals.org/doi/10.1162/jocn.2009.21283.

Bellmund, Jacob LS, Peter Gärdenfors, Edvard I Moser, and Christian F Doeller (2018). "Navigating cognition: Spatial codes for human thinking". In: *Science* 362.6415, eaat6766.

Bemis, Douglas Knox and Liina Pylkkänen (2011). "Simple Composition: A Magnetoencephalography Investigation into the Comprehension of Minimal Linguistic Phrases". In: *The Journal of Neuroscience* 31, pp. 2801–2814. URL: https://api.semanticscholar.org/CorpusID:12479545.

Boorman, Erie D, Sarah C Sweigart, and Seongmin A Park (2021). "Cognitive maps and novel inferences: a flexibility hierarchy". In: *Current Opinion in Behavioral Sciences* 38, pp. 141–149.

Brainard, David H (1997). "Psychophysics toolbox". In: *Spatial vision* 10.4, pp. 433–436.

Brennan, Jonathan, Yuval Nir, Uri Hasson, Rafael Malach, David J Heeger, and Liina Pylkkänen (Feb. 2012). "Syntactic structure building in the anterior temporal lobe during natural story listening". In: *Brain and Language* 120.2, pp. 163–173. ISSN: 0093934X. DOI: 10.1016/j.bandl.2010.04.002. URL: http://www.ncbi.nlm.nih.gov/pubmed/20472279%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2947556%20https://linkinghub.elsevier.com/retrieve/pii/S0093934X10000830.

Brennan, Jonathan R and Andrea E Martin (2020). "Phase synchronization varies systematically with linguistic structure composition". In: *Philosophical Transactions of the Royal Society B* 375.1791, p. 20190305.

Brennan, Jonathan R. and John T Hale (Jan. 2019). "Hierarchical structure guides rapid linguistic predictions during naturalistic listening". In: *PLOS ONE* 14.1. Ed. by Johan J Bolhuis, e0207741. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0207741. URL: http://dx.plos.org/10.1371/journal.pone.0207741.

Brodbeck, Christian, Alessandro Presacco, and Jonathan Z. Simon (May 2018). "Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension". In: *NeuroImage* 172, pp. 162–174. ISSN: 1053-8119. DOI: 10.1016/J.NEUROIMAGE.2018.01.042. URL: https://www.sciencedirect.com/science/article/pii/S1053811918300429?via%3Dihub.

Broderick, Michael P., Andrew J. Anderson, Giovanni M. Di Liberto, Michael J. Crosse, and Edmund C. Lalor (Mar. 2018). "Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech". In: *Current Biology* 28.5, 803–809.e3. ISSN: 09609822. DOI: 10.1016/j.cub.2018.01.080. URL: https://linkinghub.elsevier.com/retrieve/pii/S0960982218301465.

Canolty, Ryan T et al. (2006). "High gamma power is phase-locked to theta oscillations in human neocortex". In: *science* 313.5793, pp. 1626–1628.

Chomsky, Noam (June 2013). "Problems of projection". In: *Lingua* 130, pp. 33–49. ISSN: 00243841. DOI: 10.1016/j.lingua.2012.12.003. URL: https://www.sciencedirect.com/science/article/pii/S002438411200277X%20https://linkinghub.elsevier.com/retrieve/pii/S002438411200277X.

– (2014). *Aspects of the Theory of Syntax*. 11. MIT press.

Coopmans, Cas W, Helen De Hoop, Peter Hagoort, and Andrea E Martin (2022). "Effects of structure and meaning on cortical tracking of linguistic units in naturalistic speech". In: *Neurobiology of Language* 3.3, pp. 386–412.

Coopmans, Cas W, Helen De Hoop, Karthikeya Kaushik, Peter Hagoort, and Andrea E Martin (2022). "Hierarchy in language interpretation: evidence from behavioural experiments and computational modelling". In: *Language, Cognition and Neuroscience* 37.4, pp. 420–439.

Coopmans, Cas W, Karthikeya Kaushik, and Andrea E Martin (2023). "Hierarchical structure in language and action: A formal comparison." In: *Psychological Review*.

Crosse, Michael J., Giovanni M. Di Liberto, Adam Bednar, and Edmund C. Lalor (2016). "The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli". In: *Frontiers in Human Neuroscience* 10.NOV2016. Cited by: 254; All Open Access, Gold Open Access, Green Open Access. DOI: 10.3389/fnhum.2016.00604. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85007454294&doi=10.3389%2ffnhum.2016.00604&partnerID=40&md5=e37279263dc2543f987cc28a359a725a.

Deoisres, Suwijak, Yuhan Lu, Frederique J. Vanheusden, Steven L. Bell, and David M. Simpson (July 2023). "Continuous speech with pauses inserted between words increases cortical tracking of speech envelope". In: *PLOS ONE* 18.7, pp. 1–21. DOI: 10.1371/journal.pone.0289288. URL: https://doi.org/10.1371/journal.pone.0289288.

Di Liberto, Giovanni M., Michael J. Crosse, and Edmund C. Lalor (Mar. 2018). "Cortical measures of phoneme-level speech encoding correlate with the perceived clarity of natural speech". In: *eNeuro* 5.2, ENEURO.0084–18.2018. ISSN: 23732822. DOI: 10.1523/ENEURO.0084-18.2018. URL: http://eneuro.org/lookup/doi/10.1523/ENEURO.0084-18.2018.

Di Liberto, Giovanni M., James A. O'Sullivan, et al. (Sept. 2015). "Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing". In: *Current Biology* 25.19, pp. 2457–2465. ISSN: 09609822. DOI: 10.1016/j.cub.2015.08.030. URL: http://www.sciencedirect.com/science/article/pii/S0960982215010015%20http://www.ncbi.nlm.nih.gov/pubmed/26412129%20https://linkinghub.elsevier.com/retrieve/pii/S0960982215010015.

Ding, Nai, Lucia Melloni, Hang Zhang, Xing Tian, and David Poeppel (Jan. 2016). "Cortical tracking of hierarchical linguistic structures in connected speech". In: *Nature Neuroscience* 19.1, pp. 158–164. ISSN: 1097-6256. DOI: 10.1038/nn.4186. URL: http://www.nature.com/articles/nn.4186.

Ding, Nai, Xunyi Pan, Cheng Luo, Naifei Su, Wen Zhang, and Jianfeng Zhang (Jan. 2018). "Attention is required for knowledge-based sequential grouping: Insights from the integration of syllables into words". In: *Journal of Neuroscience* 38.5, pp. 1178–1188. ISSN: 15292401. DOI: 10.1523/JNEUROSCI.2606-17.2017.

Doelling, Keith B, Luc H Arnal, Oded Ghitza, and David Poeppel (2014). "Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing". In: *Neuroimage* 85, pp. 761–768.

Doelling, Keith B., M. Florencia Assaneo, Dana Bevilacqua, Bijan Pesaran, and David Poeppel (2019). "An oscillator model better predicts cortical entrainment to music". In: *Proceedings of the National Academy of Sciences*. ISSN: 0027-8424. DOI: 10.1073/pnas.1816414116.

Donhauser, Peter W and Sylvain Baillet (2019). "Two Distinct Neural Timescales for Predictive Speech Processing." In: *Neuron*. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2019.10.019.

– (Jan. 2020). "Two Distinct Neural Timescales for Predictive Speech Processing". In: *Neuron* 105.2, 385–393.e9. ISSN: 08966273. DOI: 10.1016/j.neuron.2019.10.019. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627319308931%20https://www.sciencedirect.com/science/article/abs/pii/S0896627319308931%20https://www.sciencedirect.com/science/article/pii/S0896627319308931%20https://doi.org/10.1016/j.neuron.2019.10.019.

Doumas, Leonidas AA, John E Hummel, and Catherine M Sandhofer (2008). "A theory of the discovery and predication of relational concepts." In: *Psychological review* 115.1, p. 1.

Elman, J (June 1990). "Finding structure in time". In: *Cognitive Science* 14.2, pp. 179–211. ISSN: 03640213. DOI: 10.1016/0364-0213(90)90002-E. URL: https://www.sciencedirect.com/science/article/pii/036402139090002E%20http://doi.wiley.com/10.1016/0364-0213(90)90002-E.

Elman, Jeffrey L (1991). "Distributed representations, simple recurrent networks, and grammatical structure". In: *Machine learning* 7, pp. 195–225.

Etard, Octave and Tobias Reichenbach (July 2019). "Neural Speech Tracking in the Theta and in the Delta Frequency Band Differentially Encode Clarity and Comprehension of Speech in Noise." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 39.29, pp. 5750–5759. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.1828-18.2019. URL: http://www.ncbi.nlm.nih.gov/pubmed/31109963%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6636082.

Federmeier, Kara D., Edward W. Wlotko, Esmeralda De Ochoa-Dewald, and Marta Kutas (2007). "Multiple effects of sentential constraint on word processing". In: *Brain Research* 1146, pp. 75–84.

Forte, Antonio Elia, Octave Etard, and Tobias Reichenbach (Oct. 2017). "The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention". In: *eLife* 6. Ed. by Barbara G Shinn-Cunningham, e27203. ISSN: 2050-084X. DOI: 10.7554/eLife.27203. URL: https://doi.org/10.7554/eLife.27203.

Frank, Stefan L., Leun J. Otten, Giulia Galli, and Gabriella Vigliocco (Jan. 2015). "The ERP response to the amount of information conveyed by words in sentences". In: *Brain and Language* 140, pp. 1–11. ISSN: 0093-934X. DOI: 10.1016/J.BANDL.2014.10.006. URL: https://www.sciencedirect.com/science/article/pii/S0093934X14001515?via%7B%5C%%7D3Dihub.

Frank, Stefan L. and Roel M. Willems (Oct. 2017). "Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension". In: *Language, Cognition and Neuroscience* 32.9, pp. 1192–1203. ISSN: 2327-3798. DOI: 10.1080/23273798.2017.1323109. URL: https://www.tandfonline.com/doi/full/10.1080/23273798.2017.1323109.

Frank, Stefan L. and Jinbiao Yang (May 2018). "Lexical representation explains cortical entrainment during speech comprehension". In: *PLOS ONE* 13.5. Ed. by Jonathan R. Brennan, e0197304. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0197304. arXiv: 1706.05656. URL: https://dx.plos.org/10.1371/journal.pone.0197304.

Friston, Karl (Feb. 2010). "The free-energy principle: A unified brain theory?" In: 11.2, pp. 127–138. ISSN: 1471003X. DOI: 10.1038/nrn2787.

– (Aug. 2015). "Active inference and free energy". In: *NeuroImage* 106, pp. 299–312. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2015.04.009. URL: http://linkinghub.elsevier.com/retrieve/pii/S1053811915002643%20http://www.ncbi.nlm.nih.gov/pubmed/25899227%20http://www.sciencedirect.com/science/article/pii/S1053811915002643%20http://www.sciencedirect.com/science/article/pii/S1053811915002643/pdfft?md5=2e1e1e1e1e1e1e1e1e1e1e1e1e1e1e1e1e1%7B%5C&%7Dpid=1-s2.0-S1053811915002643-main.pdf.

Giraud, Anne-Lise and Luc H Arnal (Dec. 2018). "Hierarchical Predictive Information Is Channeled by Asymmetric Oscillatory Activity." In: *Neuron* 100.5, pp. 1022–1024. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2018.11.020. URL: http://www.ncbi.nlm.nih.gov/pubmed/30521776.

Giraud, Anne-Lise and David Poeppel (Apr. 2012). "Cortical oscillations and speech processing: emerging computational principles and operations". In: *Nature Neuroscience* 15.4, pp. 511–517. ISSN: 1546-1726. DOI: 10.1038/nn.3063. URL: http://www.ncbi.nlm.nih.gov/pubmed/22426255%20http://www.nature.com/articles/nn.3063.

Gramfort, Alexandre et al. (2013). "MEG and EEG Data Analysis with MNE-Python". In: *Frontiers in Neuroscience* 7.267, pp. 1–13. DOI: 10.3389/fnins.2013.00267.

Guest, Olivia and Andrea E Martin (2023). "On logical inference over brains, behaviour, and artificial neural networks". In: *Computational Brain & Behavior*, pp. 1–15.

Haegens, Saskia and Elana Zion Golumbic (2018). "Rhythmic facilitation of sensory processing: A critical review". In: *Neuroscience & Biobehavioral Reviews* 86, pp. 150–165.

Hale, John T, Luca Campanelli, Jixing Li, Shohini Bhattasali, Christophe Pallier, and Jonathan R Brennan (2022). "Neurocomputational models of language processing". In: *Annual Review of Linguistics* 8, pp. 427–446.

Haufe, Stefan et al. (Feb. 2014). "On the interpretation of weight vectors of linear models in multivariate neuroimaging". In: *NeuroImage* 87, pp. 96–110. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.10.067. URL: http://linkinghub.elsevier.com/retrieve/pii/S1053811913010914%20http://www.sciencedirect.com/science/article/pii/S1053811913010914.

Heilbron, Micha, Kristijan Armeni, Jan-Mathijs Schoffelen, Peter Hagoort, and Floris P De Lange (2022). "A hierarchy of linguistic predictions during natural language comprehension". In: *Proceedings of the National Academy of Sciences* 119.32, e2201968119.

Jain, Shailee, Vy A Vo, Leila Wehbe, and Alexander G Huth (2023). "Computational language modeling and the promise of in silico experimentation". In: *Neurobiology of Language*, pp. 1–65.

Jensen, Ole and John E Lisman (2000). "Position reconstruction from an ensemble of hippocampal place cells: contribution of theta phase coding". In: *Journal of neurophysiology* 83.5, pp. 2602–2609.

Kaufeld, Greta, Hans Rutger Bosker, Sanne Ten Oever, Phillip M Alday, Antje S Meyer, and Andrea E Martin (2020). "Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy". In: *Journal of Neuroscience* 40.49, pp. 9467–9475.

Kaufeld, Greta, Wibke Naumann, Antje S Meyer, Hans Rutger Bosker, and Andrea E Martin (2020). "Contextual speech rate influences morphosyntactic prediction and integration". In: *Language, Cognition and Neuroscience* 35.7, pp. 933–948.

Keitel, Anne, Robin A A Ince, Joachim Gross, and Christoph Kayser (2016a). "Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks". In: *NeuroImage* 147, pp. 32–42. DOI: 10.1016/j.neuroimage.2016.11.062. URL: http://eprints.gla.ac.uk/132017/1/132017.pdf.

– (2016b). "Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks". In: *NeuroImage* 147, pp. 32–42. DOI: 10.1016/j.neuroimage.2016.11.062. URL: http://eprints.gla.ac.uk/132017/1/132017.pdf.

Keller, Georg B and Thomas D Mrsic-Flogel (2018). "Predictive processing: a canonical cortical computation". In: *Neuron* 100.2, pp. 424–435.

Kleiner, Mario, David Brainard, and Denis Pelli (2007). "What's new in Psychtoolbox-3?" In.

Koelsch, Stefan, Peter Vuust, and Karl Friston (2019). "Predictive Processes and the Peculiar Case of Music". In: *Trends in Cognitive Sciences* 23.1. ISSN: 1879307X. DOI: 10.1016/j.tics.2018.10.006.

Kösem, Anne, Virginie van Wassenhove, and Virginie van Wassenhove (2017). "Distinct contributions of low- and high-frequency neural oscillations to speech comprehension". In: *Language, Cognition and Neuroscience* 32.5, pp. 536–544. ISSN: 23273801. DOI: 10.1080/23273798.2016.1238495. URL: https://www.tandfonline.com/doi/full/10.1080/23273798.2016.1238495.

Kuperberg, Gina R. (May 2007). "Neural mechanisms of language comprehension: Challenges to syntax". In: *Brain Research* 1146.1, pp. 23–49. ISSN: 00068993. DOI: 10.1016/j.brainres.2006.12.063.

Lakatos, Peter, George Karmos, Ashesh D Mehta, Istvan Ulbert, and Charles E Schroeder (Apr. 2008). "Entrainment of neuronal oscillations as a mechanism of attentional selection." In: *Science (New York, N.Y.)* 320.5872, pp. 110–3. ISSN: 1095-9203. DOI: 10.1126/science.1154735. URL: http://www.ncbi.nlm.nih.gov/pubmed/18388295.

Levy, Roger (2008). "Expectation-based syntactic comprehension". In: *Cognition* 106.3, pp. 1126–1177.

Lewis, Ashley G., Jan-Mathijs Schoffelen, Herbert Schriefers, and Marcel Bastiaansen (2016). "A Predictive Coding Perspective on Beta Oscillations during Sentence-Level Language Comprehension". In: *Frontiers in Human Neuroscience* 10. ISSN: 1662-5161. DOI: 10.3389/fnhum.2016.00085. URL: https://www.frontiersin.org/articles/10.3389/fnhum.2016.00085.

Lewis, Ashley Glen, Jan-Mathijs Schoffelen, Christian Hoffmann, Marcel Bastiaansen, and Herbert Schriefers (May 2017). "Discourse-level semantic coherence influences beta oscillatory dynamics and the N400 during sentence comprehension". In: *Language, Cognition and Neuroscience* 32.5, pp. 601–617. ISSN: 2327-3798. DOI: 10.1080/23273798.2016.1211300. URL: https://www.tandfonline.com/doi/full/10.1080/23273798.2016.1211300.

Lewis, Ashley Glen, Lin Wang, and Marcel Bastiaansen (Sept. 2015). "Fast oscillatory dynamics during language comprehension: Unification versus maintenance and prediction?" In: *Brain and Language* 148, pp. 51–63. DOI: 10.1016/j.bandl.2015.01.003. URL: http://linkinghub.elsevier.com/retrieve/pii/S0093934X15000048.

Lisman, John E and Ole Jensen (2013). "The theta-gamma neural code". In: *Neuron* 77.6, pp. 1002–1016.

MacDonald, Maryellen C, Neal J Pearlmutter, and Mark S Seidenberg (1994). "The lexical nature of syntactic ambiguity resolution." In: *Psychological review* 101.4, p. 676.

Mahowald, Kyle, Anna A. Ivanova, Idan A. Blank, Nancy Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko (2023). *Dissociating language and thought in large language models: a cognitive perspective*. arXiv: 2301.06627 [cs.CL].

Manning, Christopher D, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J Bethard, and David McClosky (2014). "The Stanford CoreNLP natural language processing toolkit". In: *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pp. 55–60.

Marcus, Gary F (1998). "Rethinking eliminative connectionism". In: *Cognitive psychology* 37.3, pp. 243–282.

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

Martin, Andrea E and Leonidas AA Doumas (2017). "A mechanism for the cortical computation of hierarchical linguistic structure". In: *PLoS biology* 15.3, e2000663.

– (2019). "Predicate learning in neural systems: using oscillations to discover latent structure". In: *Current Opinion in Behavioral Sciences* 29, pp. 77–83.

– (2020). "Tensors and compositionality in neural systems". In: *Philosophical Transactions of the Royal Society B* 375.1791, p. 20190306.

Martin, Andrea E, Philip J Monahan, and Arthur G Samuel (2017). "Prediction of agreement and phonetic overlap shape sublexical identification". In: *Language and speech* 60.3, pp. 356–376.

Martin, Andrea E. (Aug. 2020). "A Compositional Neural Architecture for Language". In: *Journal of Cognitive Neuroscience* 32.8, pp. 1407–1427. ISSN: 0898-929X. DOI: 10.1162/jocn_a_01552. URL: https://doi.org/10.1162/jocn_a_01552%20https://www.mitpressjournals.org/doi/abs/10.1162/jocn_a_01552.

Martin, Andrea E. and B. McElree (2011). "Direct-access retrieval during sentence comprehension: Evidence from Sluicing." In: *Journal of memory and language* 64, pp. 327–343. DOI: 10.1016/j.jml.2010.12.006.

McCoy, R. Thomas, Paul Smolensky, Tal Linzen, Jianfeng Gao, and Asli Celikyilmaz (June 2023). "How Much Do Language Models Copy From Their Training Data? Evaluating Linguistic Novelty in Text Generation Using RAVEN". In: *Transactions of the Association for Computational Linguistics* 11, pp. 652–670. ISSN: 2307-387X. DOI: 10.1162/tacl_a_00567. eprint: https://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00567/2141008/tacl_a_00567.pdf. URL: https://doi.org/10.1162/tacl%5C_a%5C_00567.

Meyer, Lars (Oct. 2018). "The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms". In: *European Journal of Neuroscience* 48.7, pp. 2609–2621. ISSN: 0953816X. DOI: 10.1111/ejn.13748. URL: http://doi.wiley.com/10.1111/ejn.13748.

Meyer, Lars and Matthias Gumbert (Aug. 2018a). "Synchronization of Electrophysiological Responses with Speech Benefits Syntactic Information Processing". In: *Journal of Cognitive Neuroscience* 30.8, pp. 1066–1074. ISSN: 0898-929X. DOI: 10.1162/jocn_a_01236. URL: https://www.mitpressjournals.org/doi/abs/10.1162/jocn%7B%5C_%7Da%7B%5C_%7D01236.

Meyer, Lars and Matthias Gumbert (2018b). "Synchronization of electrophysiological responses with speech benefits syntactic information processing". In: *Journal of cognitive neuroscience* 30.8, pp. 1066–1074.

Meyer, Lars, Yue Sun, and Andrea E Martin (2020a). ""Entraining" to speech, generating language?" In: *Language, Cognition and Neuroscience* 35.9, pp. 1138–1148.

– (2020b). "Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing". In: *Language, Cognition and Neuroscience* 35.9, pp. 1089–1099.

Molinaro, Nicola, Horacio A. Barber, and Manuel Carreiras (2011). "Grammatical agreement processing in reading: ERP findings and future directions". In: *Cortex* 47.8, pp. 908–930. ISSN: 0010-9452. DOI: https://doi.org/10.1016/j.cortex.2011.02.019. URL: http://www.sciencedirect.com/science/article/pii/S0010945211000591.

Nelson, Matthew J et al. (Apr. 2017). "Neurophysiological dynamics of phrase-structure building during sentence processing." In: *Proceedings of the National Academy of Sciences of the United States of America*, p. 201701590. ISSN: 1091-6490. DOI: 10.1073/pnas.1701590114. URL: http://www.ncbi.nlm.nih.gov/pubmed/28416691.

Nieuwland, Mante S (2019). "Do 'early'brain responses reveal word form prediction during language comprehension? A critical review". In: *Neuroscience & Biobehavioral Reviews* 96, pp. 367–400.

Nieuwland, Mante S et al. (2018). "Large-scale replication study reveals a limit on probabilistic prediction in language comprehension". In: *ELife* 7, e33468.

Nieuwland, Mante S. et al. (Feb. 2020). "Dissociable effects of prediction and integration during language comprehension: evidence from a large-scale study using brain potentials". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1791, p. 20180522. ISSN: 0962-8436. DOI: 10.1098/rstb.2018.0522. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2018.0522.

Obleser, Jonas and Christoph Kayser (Sept. 2019a). "Neural Entrainment and Attentional Selection in the Listening Brain". In: *Trends in Cognitive Sciences* 23.11, pp. 913–926. ISSN: 1364-6613. DOI: 10.1016/J.TICS.2019.08.004. URL: https://www.sciencedirect.com/science/article/pii/S1364661319302050%7B%5C#%7Ds0005.

– (2019b). "Neural entrainment and attentional selection in the listening brain". In: *Trends in cognitive sciences* 23.11, pp. 913–926.

Oganian, Yulia, Katsuaki Kojima, Assaf Breska, Chang Cai, Anne Findlay, Edward Chang, et al. (2023). "Phase alignment of low-frequency neural activity to the amplitude envelope of speech reflects evoked responses to acoustic edges, not oscillatory entrainment". In: *Journal of Neuroscience* 43.21, pp. 3909–3921.

Oganian, Yulia, Katsuaki Kojima, Assaf Breska, Chang Cai, Anne Findlay, Edward F. Chang, et al. (2023). "Phase Alignment of Low-Frequency Neural Activity to the Amplitude Envelope of Speech Reflects Evoked Responses to Acoustic Edges, Not Oscillatory Entrainment". In: *Journal of Neuroscience* 43.21, pp. 3909–3921. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.1663-22.2023. eprint: https://www.jneurosci.org/content/43/21/3909.full.pdf. URL: https://www.jneurosci.org/content/43/21/3909.

Oostenveld, Robert, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen (2011). "Cluster-based permutation tests in neuroscience: theory, examples, and pitfalls". In: *NeuroImage* 56.3, pp. 908–916.

Peelle, Jonathan E. and Matthew H. Davis (Jan. 2012). "Neural oscillations carry speech rhythm through to comprehension". In: *Frontiers in Psychology* 3.SEP, p. 320. ISSN: 16641078. DOI: 10.3389/fpsyg.2012.00320. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3434440%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract.

Pittman-Polletta, Ben R. and Bernat Kocsis (2022). "Assessing Neural Circuit Interactions and Dynamics with Phase-Amplitude Coupling". In: *Electrophysiological Recording Techniques*. Ed. by Robert P. Vertes and Timothy Allen. New York, NY: Springer US, pp. 125–146. ISBN: 978-1-0716-2631-3. DOI: 10.1007/978-1-0716-2631-3_6. URL: https://doi.org/10.1007/978-1-0716-2631-3_6.

Pylkkänen, Liina (2020). "Neural basis of basic composition: what we have learned from the red–boat studies and their extensions". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1791, p. 20190299. DOI: 10.1098/rstb.2019.0299. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2019.0299. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2019.0299.

Pylkkänen, Liina and Jonathan R. Brennan (2019). "The neurobiology of syntactic and semantic structure building". In: *The Cognitive Neurosciences*. Ed. by M. Gazzaniga, G. Mangun, and D. Poeppel. 6th ed.

Radford, Alec, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever (2019). "Language models are unsupervised multitask learners". In: *OpenAI Blog* 1.8, p. 9.

Rao, Rajesh P N and Dana H. Ballard (Jan. 1999). "Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects". In: *Nature Neuroscience* 2.1, pp. 79–87. ISSN: 10976256. DOI: 10.1038/4580.

Saffran, Jenny R, Richard N Aslin, and Elissa L Newport (1996). "Statistical learning by 8-month-old infants". In: *Science* 274.5294, pp. 1926–1928.

Slaats, Sophie and Andrea E Martin (2023). "What's surprising about surprisal". In.

Slaats, Sophie, Hugo Weissbart, Jan-Mathijs Schoffelen, Antje S Meyer, and Andrea E Martin (2023). "Delta-band neural responses to individual words are modulated by sentence processing". In: *Journal of Neuroscience*.

Spitzer, Bernhard and Saskia Haegens (2017). "Beyond the status quo: a role for beta oscillations in endogenous content (re) activation". In: *eneuro* 4.4.

Ten Oever, Sanne, Sara Carta, Greta Kaufeld, and Andrea E Martin (2022). "Neural tracking of phrases in spoken language comprehension is automatic and task-dependent". In: *Elife* 11, e77468.

Ten Oever, Sanne, Karthikeya Kaushik, and Andrea E Martin (2022). "Inferring the nature of linguistic computations in the brain". In: *PLoS Computational Biology* 18.7, e1010269.

Ten Oever, Sanne and Andrea E Martin (2021). "An oscillating computational model can track pseudo-rhythmic speech by using linguistic predictions". In: *Elife* 10, e68066.

Ten Oever, Sanne, Tobias Meierdierks, Felix Duecker, Tom A De Graaf, and Alexander T Sack (2020). "Phase-coded oscillatory ordering promotes the separation of closely matched representations to optimize perceptual discrimination". In: *Iscience* 23.7.

Ten Oever, Sanne and Alexander T Sack (2015). "Oscillatory phase shapes syllable perception". In: *Proceedings of the National Academy of Sciences* 112.52, pp. 15833–15837.

Teng, Xiangbin, Xing Tian, Keith Doelling, and David Poeppel (Oct. 2018). "Theta band oscillations reflect more than entrainment: behavioral and neural evidence demonstrates an active chunking process". In: *European Journal of Neuroscience* 48.8, pp. 2770–2782. ISSN: 0953816X. DOI: 10.1111/ejn.13742. URL: http://edmond.mpdl.mpg.de/imeji/collection/kZalRMtxa19mlRyG%20http://doi.wiley.com/10.1111/ejn.13742.

Tezcan, Filiz, Hugo Weissbart, and Andrea E Martin (July 2023). "A tradeoff between acoustic and linguistic feature encoding in spoken language comprehension". In: *eLife* 12. Ed. by Jonathan Erik Peelle and Andrew J King, e82386. ISSN: 2050-084X. DOI: 10.7554/eLife.82386. URL: https://doi.org/10.7554/eLife.82386.

Van Schijndel, Marten and Tal Linzen (2021). "Single-stage prediction models do not explain the magnitude of syntactic disambiguation difficulty". In: *Cognitive science* 45.6, e12988.

Van Veen, B.D., W. Van Drongelen, M. Yuchtman, and A. Suzuki (1997). "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering". In: *IEEE Transactions on Biomedical Engineering* 44.9, pp. 867–880. DOI: 10.1109/10.623056.

Von der Malsburg, Christoph (1995). "Binding in models of perception and brain function". In: *Current opinion in neurobiology* 5.4, pp. 520–526.

– (1999). "The what and why of binding: the modeler's perspective". In: *Neuron* 24.1, pp. 95–104.

Weissbart, Hugo, Katerina D. Kandylaki, and Tobias Reichenbach (Jan. 2020). "Cortical Tracking of Surprisal during Continuous Speech Comprehension". In: *Journal of Cognitive Neuroscience* 32.1, pp. 155–166. ISSN: 0898-929X. DOI: 10.1162/jocn_a_01467. URL: https://www.mitpressjournals.org/doi/abs/10.1162/jocn%7B%5C_%7Da%7B%5C_%7D01467.

Zaccarella, Emiliano, Marianne Schell, and Angela D Friederici (2017). "Reviewing the functional basis of the syntactic Merge mechanism for language: A coordinate-based activation likelihood estimation meta-analysis". In: DOI: 10.1016/j.neubiorev.2017.06.011. URL: http://dx.doi.org/10.1016/j.neubiorev.2017.06.011.

Zioga, Ioanna, Hugo Weissbart, Ashley G. Lewis, Saskia Haegens, and Andrea E. Martin (2023). "Naturalistic Spoken Language Comprehension Is Supported by Alpha and Beta Oscillations". In: *Journal of Neuroscience* 43.20, pp. 3718–3732. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.1500-22.2023. eprint: https://www.jneurosci.org/content/43/20/3718.full.pdf. URL: https://www.jneurosci.org/content/43/20/3718.

Zoefel, Benedikt, Sanne ten Oever, and Alexander T. Sack (2018). "The involvement of endogenous neural oscillations in the processing of rhythmic input: More than a regular repetition of evoked neural responses". In: *Frontiers in Neuroscience* 12.MAR, pp. 1–13. ISSN: 1662453X. DOI: 10.3389/fnins.2018.00095.

Zou, Jiajie et al. (2021). "θ-Band Cortical Tracking of the Speech Envelope Shows the Linear Phase Property". In: *eNeuro* 8.4. DOI: 10.1523/ENEURO.0058-21.2021. eprint: https://www.eneuro.org/content/8/4/ENEURO.0058-21.2021.full.pdf. URL: https://www.eneuro.org/content/8/4/ENEURO.0058-21.2021.

Zuur, Alain F., Elena N. Ieno, and Chris S. Elphick (2010). "A protocol for data exploration to avoid common statistical problems". In: *Methods in Ecology and Evolution* 1.1, pp. 3–14. DOI: https://doi.org/10.1111/j.2041-210X.2009.00001.x. eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2041-210X.2009.00001.x. URL: https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.2041-210X.2009.00001.x.