Event-related causality in stereo-EEG discriminates syntactic processing of noun phrases and verb phrases

Journal of Neural Engineering

PAPER

Event-related causality in stereo-EEG discriminates syntactic processing of noun phrases and verb phrases

Andrea Cometa[1,2], **Piergiorgio d'Orio**[3,4], **Martina Revay**[3,4], **Franco Bottoni**[5], **Claudia Repetto**[6], **Giorgio Lo Russo**[3], **Stefano F Cappa**[7,10], **Andrea Moro**[2], **Silvestro Micera**[1,8,11] and **Fiorenzo Artoni**[1,8,9,11,*]

[1] The BioRobotics Institute and Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, Viale Rinaldo Piaggio 34, Pontedera 56025, Italy
[2] Scuola Universitaria Superiore IUSS, Piazza della Vittoria 15, Pavia 27100, Italy
[3] "Claudio Munari" Epilepsy Surgery Centre, Azienda Socio-Sanitaria Territoriale Grande Ospedale Metropolitano Niguarda, Piazza dell'Ospedale Maggiore 3, Milano 20162, Italy
[4] Department of Medicine and Surgery, Unit of Neuroscience, Universita' di Parma, Via Giovanni Battista Grassi 74, Parma, Italy
[5] Istituto Clinico Humanitas, IRCCS, Via Alessandro Manzoni 56, Rozzano 20089, Italy
[6] Department of Psychology, Università Cattolica del Sacro Cuore, Largo A. Gemelli 1, Milan 20123, Italy
[7] IRCCS Mondino Foundation National Institute of Neurology, Via Mondino 2, Pavia 27100, Italy
[8] Bertarelli Foundation Chair in Translational NeuroEngineering, Center for Neuroprosthetics and School of Engineering, Ecole Polytechnique Federale de Lausanne, Campus Biotech, Chemin des Mines 9, Geneva, GE CH 1202, Switzerland
[9] Department of Basic Neurosciences, University of Geneva, Campus Biotech, Chemin des Mines 9, Geneva, GE CH 1202, Switzerland
[10] Cognitive Neuroscience (ICoN) Center, Scuola Universitaria Superiore IUSS, Piazza Vittoria 15, Pavia 27100, Italy
[11] These authors contributed equally to the work.
[*] Author to whom any correspondence should be addressed.

E-mail: fiorenzo.artoni@unige.ch

Keywords: SEEG, syntax, partial directed coherence, event-related causality, connectivity, speech, decoding

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
10 October 2022

REVISED
20 February 2023

ACCEPTED FOR PUBLICATION
5 April 2023

PUBLISHED
2 May 2023

Abstract

Objective. Syntax involves complex neurobiological mechanisms, which are difficult to disentangle for multiple reasons. Using a protocol able to separate syntactic information from sound information we investigated the neural causal connections evoked by the processing of homophonous phrases, i.e. with the same acoustic information but with different syntactic content. These could be either verb phrases (VP) or noun phrases. *Approach*. We used event-related causality from stereo-electroencephalographic recordings in ten epileptic patients in multiple cortical and subcortical areas, including language areas and their homologous in the non-dominant hemisphere. The recordings were made while the subjects were listening to the homophonous phrases. *Main results.* We identified the different networks involved in the processing of these syntactic operations (faster in the dominant hemisphere) showing that VPs engage a wider cortical and subcortical network. We also present a proof-of-concept for the decoding of the syntactic category of a perceived phrase based on causality measures. *Significance*. Our findings help unravel the neural correlates of syntactic elaboration and show how a decoding based on multiple cortical and subcortical areas could contribute to the development of speech prostheses for speech impairment mitigation.

1. Introduction

Traditionally, language is analyzed in relation to four main components: the acoustic level, that is the physical medium humans naturally exploit to convey information and its articulatory–phonatory counterpart; the lexicon, which is the repertoire of words expressing predicative contents and logical instructions; syntax, the set of principles to assemble larger units (phrases) from lexical items, in a recursive potentially infinite way; semantics, an interpretative component which captures the truth value conditions for each syntactic structure. However, since the acoustic and syntactic information are

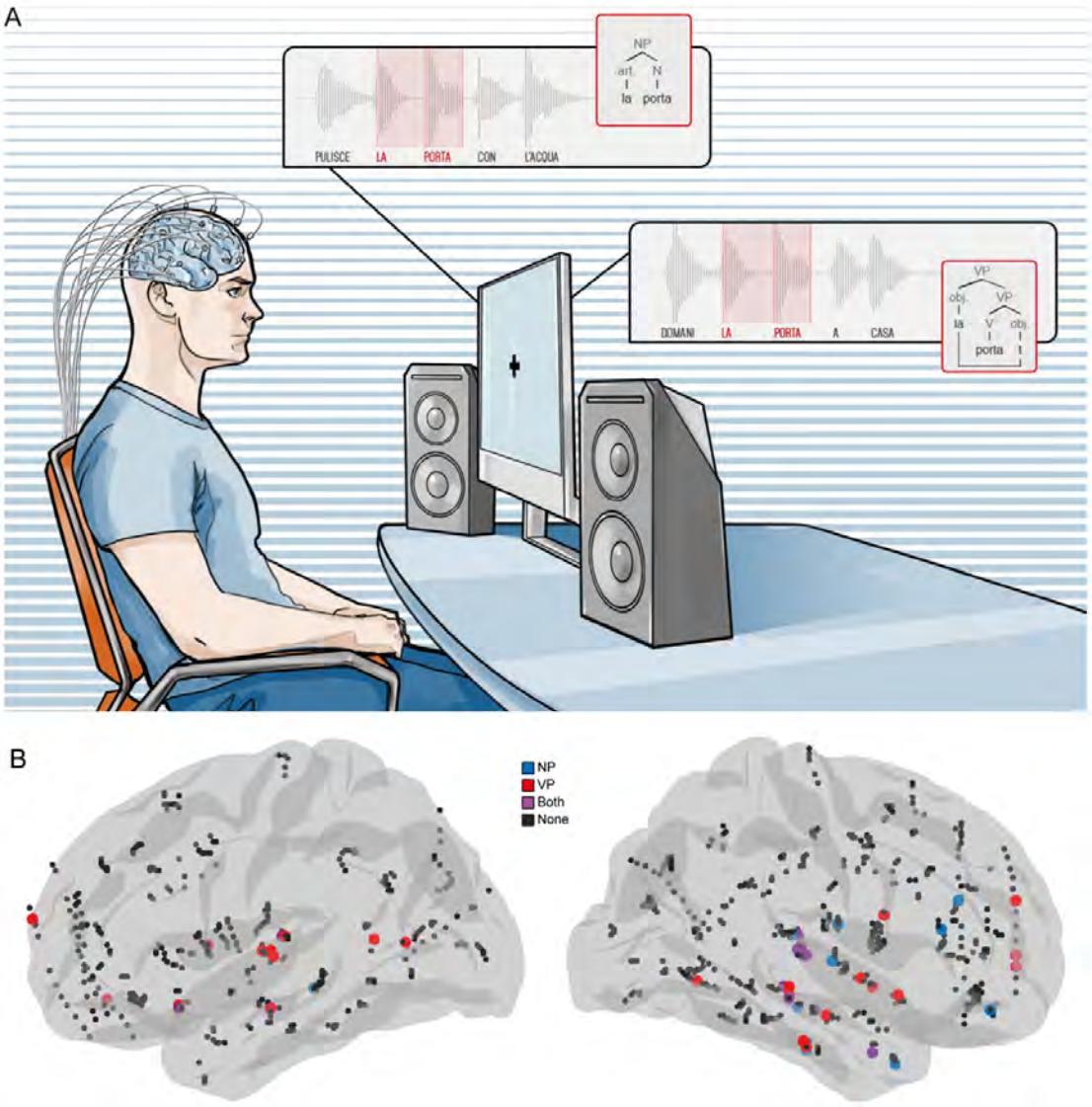

Figure 1. Experimental set-up. (A) Example of a set of homophonous sequences (i.e. strings of words with the same sound but different syntactic structure) used in the experiment. For example, in PULISCE LA PORTA CON L'ACQUA (s/he cleans the door with water), the phonemic sequence [la'pɔrta] (written here as: la porta) is a NP, while in DOMANI LA PORTA A CASA (tomorrow s/he brings her home), the same sequence is a verb phrase (VP). (B) Mini-region-of-interests (merged across all subjects) in the dominant (left) and non-dominant (right) hemispheres. Contacts involved in the NP-related network are highlighted in blue, those involved in the VP processing network are highlighted in red, and those participating in both networks are colored in purple. Adapted from Artoni *et al* (2020). CC BY 4.0.

crucially intertwined (Ding *et al* 2015), even during inner speech (Magrassi *et al* 2015, Kayne 2019), isolating syntax at the electrophysiological level appears to be an insurmountable empirical task. This is reflected in the difficulty of developing specific syntax-related tasks for experimental studies of language neurobiology and it is responsible for the relatively limited knowledge of syntax-related processing in the brain. Understanding the neural correlates of even the most basic syntactic operations, such as merging an article with a noun (N) yielding a noun phrase (NP) or a pronoun with a verb (V) yielding a verb phrase (VP) remains a crucial challenge for brain and language research (Grodzinsky and Friederici 2006).

In a recent study (Artoni *et al* 2020), we designed and used a novel protocol aimed at isolating syntactic information from the acoustic associated information by exploiting pairs of sentences containing homophonous strings (same acoustic information but completely different syntactic content). Specifically, each pair of stimuli contained the same acoustic copy of two homophonous words, which could be interpreted either as a NP or a VP (figure 1(A)). This approach was used to factor out any phonological and prosodical clue in a complete way, even at the subliminal level. We used this protocol while recording the related cortical and subcortical activation using stereo-electroencephalography (SEEG), an invasive recording technique with unparalleled signal-to-

noise ratio and recording band-width (Lachaux *et al* 2003, He *et al* 2019). Surprisingly, we found that the effect of the syntactic structure on cortical and subcortical activity was not limited to the brain areas traditionally associated with syntactic processing (i.e. Broca's area and the left posterior temporo-parietal cortex) as suggested by previous studies (Knösche *et al* 1999, Friederici *et al* 2000, Friederici and Kotz 2003, Nuñez *et al* 2011, Batterink and Neville 2013, Griffiths *et al* 2013, Weber *et al* 2016, Schell *et al* 2017, Friederici 2018, Pylkkänen 2019), but involved multiple regions in both hemispheres.

The complex mechanisms that led to these results cannot be fully described by treating the single cortical hubs as segregated structures. In fact, oscillatory neuronal activity plays an important role in organizing neurons in large-scale networks (Korzeniewska *et al* 2008) and high gamma activity arising from one brain area may induce high gamma activity in another cortical region (Engel and Singer 2001, Varela *et al* 2001, Buzsáki and Draguhn 2004). For this reason, in this study, we further exploited the SEEG signal potential to investigate the amplitude, the direction, and the specific frequencies of the interactions taking place between brain structures, that is the collection of causal links elicited by different functional situations known as effective connectivity (Penny *et al* 2004). Given the utmost importance of timing, here we analyzed the directed connectivity patterns elicited by a stimulus, i.e. the ERC. We investigated the dynamical evolution of the causal integration in response to a specific part of the time-varying stimuli (sentences)—the response window (RW)—either the NP or the VP. To reach this aim and to characterize and define the different networks involved in the processing of the syntactic operations yielding a NP or a VP we used a recently validated pipeline of ours for the evaluation of ERC in a set RW (Cometa *et al* 2021).

We also present a proof-of-concept for the decoding of the syntactic category of a perceived sentence based on causality measures which could contribute to the future development of speech prostheses for speech impairment mitigation.

2. Methods

2.1. Human subjects

In total, 23 patients were recruited. All of them underwent surgical implantations of intracerebral electrodes for refractory epilepsy (Cossu *et al* 2015) in the 'Claudio Munari' Epilepsy Surgery Center of Milan, Italy (Munari *et al* 1994, Cossu *et al* 2005). The strategy of implantation was defined purely based on clinical needs, to locate the epileptogenic zone.

All patients completed all experimental sessions. During the 24 h before the experimental recording, no seizure occurred, no alterations in the sleep/wake cycle were observed, and no additional pharmacological treatments were applied. No language or neuropsychological deficits were found in any patients. Also, no anatomical alterations were made evident by magnetic resonance. High-frequency stimulation (50 Hz, 3 mA, 5 sec) through SEEG electrodes was used to assess language dominance in all subjects. Two patients also underwent a functional magnetic resonance imaging (fMRI) study during a language task before the implantation of the electrodes.

Thirteen patients were excluded from the analysis. Eight of them exhibited pathological activity with no background rhythm in more than 50% of the SEEG contacts. The others five patients showed no implanted recording contacts with a task-related significant activation in our previous study (Artoni *et al* 2020). Full demographic data are shown in table S5.

A total of 2186 recording contacts (median 210, range 168–272) were implanted, divided into 164 electrodes (median 16.5, range 13–19). The number of contacts in the grey matter was 1439 (65.8%); 586 recording contacts in the language dominant hemisphere (DH). The DH was implanted in five subjects (median electrodes 16, range 3–18; median contacts 210, range 25–225). The non-DH (NDH) was implanted in six subjects (median electrodes 15, range 14–19; median contacts 208, range 182–272). SEEG exploration involved both hemispheres with a preference for the non-dominant side in 1 patient.

Overall, 68 electrodes were implanted in the temporal lobe (26 in DH, 42 in NDH), 43 in the frontal lobe (22 in DH, 21 in NDH), 22 in the central lobe (9 in DH, 13 in NDH), and 30 in the parieto-occipital region (9 in DH and 21 in NDH).

The present study received the approval of the Ethics Committee of ASST Grande Ospedale Metropolitano Niguarda (ID 939-2.12.2013) and informed consent was obtained from all participants.

2.2. Stimuli

The set of stimuli is based on three characteristics of Italian. First, some definite articles are pronounced exactly like some object clitic pronouns (such as [la] written as *la*; it can be both 'the—fem.sing.' or 'her—fem.sing.'). Second, the syntax of articles and clitic pronouns is very different: articles precede nouns, complements follow verbs, but object clitics are placed before the verb. Third, the Italian lexicon contains several homophonous pairs of nouns and verbs, such as ['pɔrta] (written *porta*), which can either mean 'door' or 'brings'. A set of pairs of words such as [la 'pɔrta] (written as *la porta*) can thus be interpreted either as a NP ('the door') or a VP ('brings her') depending on the syntactic context (homophonous phrases). For example, in PULISCE LA PORTA CON L'ACQUA (s/he cleans the door with water), *la porta* is a NP, while in DOMANI LA PORTA A CASA (tomorrow s/he brings her home),

la porta is a VP. We used 62 stimuli (table S1, in supplementary information), i.e. with 31 pairs of homophonous phrases.

To be sure to eliminate phonological and prosodical factors, the pronunciation of one homophonous phrase was copied in the syntactic counterpart. No other semantic or lexical distinction differentiated the two types of phrases.

The acoustic stimuli were recorded using a Sennheiser Microphone MH40P48, connected via a Firewire 400 to an Apple OSX 10.5.8 with a Motu Ultralight Mk3 sound card. The stimuli were edited and mastered using Audiodesk 3.02 and Peak Pro7, respectively. Files were generated in 16 bits, with a sampling frequency equal to 44.1 kHz; intensity was normalized to 0 Db and rendered in .wav format. All sentences were read by the same person, an Italian native speaker, male, 53 years old.

2.3. Surgical procedure and recording equipment

SEEG electrodes have a diameter of 0.8 mm. They contain 5–18 recording contacts, which are 2 mm long and spaced by 1.5 mm. The strategy of implantation was planned on 3D multimodal imaging and the electrodes were stereotactically implanted with robotic assistance. After the implantation, cone-beam computed tomography was acquired with the O-arm scanner (Medtronic) and registered to pre-implantation 3D T1-weighted MR images. Subsequently, multimodal scenes were built with the 3D Slicer software package (Fedorov *et al* 2012) and the exact position of each lead was determined both looking at multiplanar reconstructions and using the SEEG assistant tool available for 3D Slicer itself (Narizzano *et al* 2017). The spatial coordinates of each lead in individual anatomical space provided by this tool were then converted into the MNI space coordinate triplet after co-registration of the patients' space to the MNI space.

SEEG sampling rate during the experiment was set to 1 kHz (patients 1–12) or 2 kHz (patients 13–23). Recordings were carried out using a 192-channels EEG-1200 (Neurofax, Nihon Kohden). All recording contacts were re-referenced to two leads in the white matter, in which electrical stimulations did not produce any manifestation.

2.4. Recording protocol

Each subject rested in a comfortable armchair. Stimuli were delivered using the Presentation software (Neurobehavioral Systems). Phrases were delivered via audio amplifiers at the minimum volume for words to be perceived with ease, according to the subject. During stimuli delivery, subjects gazed at a 27 inches cross on a screen. A synchronization TTL trigger spike was sent to the SEEG trigger port at the beginning of the sentence. Jitter and delays were lower than 1 ms. The experiment lasted around 30 min, with no breaks. At the end of each task, subjects were always able to correctly answer short questions about the stimuli. A camera was used to control for eye movement, silence, and any unexpected behavior from the patients.

2.5. Data pre-processing

An anti-aliasing band-pass filter (0.015–500 Hz) was applied at the hardware level. Recordings acquired at 2 kHz were down-sampled to 1 kHz. Channels from which pathological activity was recorded during the task were removed by clinicians. Recordings were annotated with the events triggered by the beginning of each word in all stimulus sentences. Epochs were extracted from −1.5 s to 4.5 s time-locked to the beginning of each stimulus. The length of the epochs always ensured the inclusion of the complete stimulus presentation. Epochs with notable artifacts were rejected. Recording contacts in white matter have a lower amplitude and a narrower frequency band with respect to recording contacts in the grey matter. These visual clues were used by the clinicians to identify and exclude the contacts in white matter from subsequent analysis.

2.6. Cortico-cortical evoked potentials

During the presurgical evaluation, an effective connectivity of the implanted brain areas was assessed for each subject by evaluating the cortico-cortical evoked potentials (CCEP) elicited by single-pulse electrical stimulation (SPES) (Matsumoto *et al* 2017, Trebaul *et al* 2018, Russo *et al* 2021).

In the condition of eyes open resting wakefulness, SPES was delivered through each pair of adjacent contacts, with at 5 mA current intensity, a single pulse of 0.5 ms (biphasic rectangular stimuli of alternating polarity), at 1 Hz frequency, for 15 s.

The presence of CCEPs response following a SPES was visually verified by trained neurophysiologists.

2.7. Stimulus-evoked causality estimation

To estimate the stimulus-evoked directed connections, recording contacts were first divided into mini-regions of interest (mini-ROIs). Then, the partial directed coherence (PDC), a measure deriving from the Granger causality framework (Granger 1969, Geweke 1982, Baccalá and Sameshima 2001) was computed. Finally, a non-parametric statistical test was used to evaluate the significant connections elicited in the RW, i.e. the part of interest of the stimulus (NP or VP). This stimulus-evoked causality estimation pipeline, designed for SEEG data, is proposed in (Cometa *et al* 2021).

2.7.1. Mini-ROI extraction

Two SEEG contacts which are very close in space record almost the same signal. This could lead to artificially high causality values, which in turn (being most causality measures normalized quantities) may mask significant causality values between

distant recording contacts. Thus, for each subject, the recording contacts showing high correlation coefficients between their time series were combined into mini-ROIs. Specifically, mini-ROIs are groups of leads having an averaged across trials coefficient of determination $R^2 > 0.8$. The prototypical channel of a mini-ROI was selected as the one showing the highest linear correlation with the mini-ROI mean time series. Mini-ROIs grouping was performed independently for each subject. Most mini-ROIs were populated by just one channel, with the most numerous ones not being populated by more than 3 recording contacts. Not surprisingly, all the recording contacts assigned in a single mini-ROI were spatially very close and always belonged to the same shaft.

2.7.2. Causality estimation

Within the Granger causality framework, a time series $x_j(t)$ causes another time series $x_i(t)$ if knowledge of past samples of $x_j(t)$ reduces the prediction error for the current sample of $x_i(t)$. The relation between $x_j(t)$ and $x_i(t)$ can be estimated by fitting a time-varying multivariate autoregressive (MVAR) model on $\boldsymbol{X}(t)$:

$$\boldsymbol{X}(t) = [x_1(t), x_2(t), \ldots, x_D(t)]^T \quad (1)$$

where D is the total number of channels.

The MVAR model assumes a linear relationship between the channels in $\boldsymbol{X}(t)$ of the form:

$$\boldsymbol{X}(t) = -\sum_{k=1}^{p} A_k(t)\boldsymbol{X}(t-k) + \boldsymbol{e}(t) \quad (2)$$

where $A_k(t)$ is the time-varying DxD MVAR coefficients matrix, $\boldsymbol{e}(t)$ is a white noise process with covariance matrix W and p is the model order. The $A_k(t)$ matrices were derived by using a general linear Kalman Filter (Milde *et al* 2010). To estimate the model order p, the Bayesian information criterion was used (Schwarz 1978), resulting in $p = 4$ for all subjects.

After estimating, trial by trial, the $A_k(t)$ matrices, the single-trial time-varying $\mathrm{PDC}(f, t)$ (Astolfi *et al* 2008) was computed.

To lower the computational complexity of the pipeline, PDC time samples were down-sampled by a factor of 40 (from 6000 samples to 150). Frequencies were averaged into overlapping frequency bins (width = 50 Hz, overlap = 25 Hz, range = 0–300 Hz).

Subsequent analysis was done only in the ultra-high gamma frequency range (150–300 Hz), i.e. on frequency bins from [125–175 Hz] to [250–300 Hz].

2.7.3. Significance during the homophonous phrase

All the next steps of the algorithm were independently applied for each syntactic structure (NPs or VPs), each subject, and each frequency band f. Linear interpolation time-warping was used to align the RW across all trials (Gwin *et al* 2011, Artoni *et al* 2017, Nordin *et al* 2019, Do *et al* 2021). Baseline correction was then carried out by dividing $\mathrm{PDC}_{ij}(f, t)$, trial by trial and for each i, j $(i \neq j)$ couple independently, by its mean baseline value. The $\overline{\mathrm{PDC}_{ij}}(f, t)$ matrices were obtained by averaging $\mathrm{PDC}_{ij}(f, t)$ over trials. The mean values of the $\overline{\mathrm{PDC}_{ij}}(f, t)$ during the RW were calculated for each pair i, j $(i \neq j)$ of channels.

We subsequentially performed a statistical test aimed at identifying the strongest connections within each subject and only retaining those for subsequent analysis. We called these strongest connections.

2.7.3.1. Significant connections

The mean values of the $\overline{\mathrm{PDC}_{ij}}(f, t)$ during the RW were compared against a null distribution: to generate the null (permutation) distribution and to control for false discovery rate (Nichols and Holmes 2002, Maris and Oostenveld 2007) the time samples of the $\overline{\mathrm{PDC}_{ij}}(f, t)$ were shuffled 1000 times and the mean values during the RW were re-computed for each permutation. The maximum mean value across all channel couples was retained for each permutation. An arbitrary *significance threshold* was then set in order to detect significant connections. For each pair of recording contacts we calculated the fraction of instances in the null distribution that were greater than the mean RW causality occurring between that pair. The connection was deemed significant if this value was below the arbitrary *significance threshold*. We set the *significance threshold* to 0.33, being the lowest one that allowed the arising of at least one significant connection for either NPs or VPs in every subject, in at least one of the considered frequency bins (from [125–175 Hz] to [250–300 Hz]).

It is important to note that the null distribution can be also computed by shifting the original time series (Crowther *et al* 2019) or by randomizing their phases (Brunner *et al* 2019), prior to calculating the PDC values. However, in (Cometa *et al* 2021) we proved that these two alternatives do not bring any advantage while being more computationally cumbersome.

2.8. Inter-subject analysis

In SEEG experiments, the location of implantation of the electrodes changes drastically across subjects. It is therefore very difficult to combine the results and handle the inter-subject differences. We decided to use a *patchwork* approach: we applied all the steps used to estimate the stimulus-evoked causality (i.e. mini-ROI extraction, PDC calculation and significance assessment) independently for each subject. The resulting significant connections were combined across subjects by concatenating them, and the subsequent analysis on their emerging properties were done on the set of all significant connections arising from all subjects.

2.9. Latency analysis

To detect the peaks in connectivity during the RW of the stimuli, the average connectivity time series were

first smoothed. A Savgol filter was used (Guiñón *et al* 2007). The polynomial order was set to 2, with nine-samples long windows. The window size was chosen as the knee of the curve formed by the sum of absolute differences between the smoothed time series and the raw ones for different window lengths. The latencies were defined as the time instant at which the maximum of each smoothed time series occurred, within the homophonous phrase interval.

2.10. Cortical surface plotting

Mini-ROIs, active directed connections, and active cortical areas were graphically represented using the *BrainNet Viewer* toolbox for Matlab (Xia *et al* 2013). Plotting was done using MNI coordinates on a FreeSurfer *fsaverage* template (Fischl 2012, Wu *et al* 2018).

3. Decoding

3.1. RW prediction

The prediction of the phase of the stimulus was carried out on a trial-by-trial basis. The total number of trials, across all subjects, was 700. All the connections were used. For each subject, all the time-varying connectivity amplitudes were divided into overlapping bins of size 20 samples and step 1 sample and then averaged within each window, resulting in one value per subject per time window. These values were fed to a long short-term memory network (LSTM) (Hochreiter and Schmidhuber 1997) together with the labels corresponding to the stimulus phase (baseline, sentence start, RW, sentence ending) of the last sample of the corresponding overlapping window. We used a LSTM instead of a simpler approach such as the linear regression to not be bound by the assumptions of independence, normality and homoscedasticity and for the ability of the LSTM to exploit the temporal structure of the input—i.e. the time-varying PDC values—to make a prediction.

The training was carried out using a leave-one-subject-out (LOSO) cross-validation procedure. For each iteration of the LOSO cross-validation procedure, the time-varying connectivity amplitudes for one subject were held out to be used as the test set, while the PDC values of the other nine subjects were used as the training set. Two trials were removed from the training set and used as the validation set. The decoder hyperparameters were optimized according to the performance on the validation set. Hyperparameter optimization was performed using a grid-search on [0.00001, 0.0001, 0.001, 0.01] for the learning rate, [16, 32, 64, 128] for the number of hidden units of the LSTM, and [0, 0.1, 0.3, 0.5] for the dropout (i.e. the fraction of weights that are randomly forgot after each training epoch, used to avoid overfitting).

The resulting best hyperparameters were used to train the LSTM on the training set. The training procedure was stopped after 100 epochs. The accuracy pertaining to each fold was calculated on the held out test set. The final accuracy was obtained by averaging the accuracies across all folds of the LOSO cross-validation.

A weighted version of the categorical cross-entropy (Abadi *et al* 2015, Ho and Wookey 2020) was used as the loss function to minimize during the training of the LSTM, with the weights for each class inversely proportional to the length of the stimulus phase.

Code implementation was based on the *TensorFlow* package for *Python* (Abadi *et al* 2015).

3.2. Syntactic content decoding

The prediction of the content of the homophonous phrases (NP vs VP) was carried out on a trial-by-trial basis. Only the significant connections were selected, regardless of whether the connections were significant during NPs or VPs processing. For each time point, a number of values equal to the number of significant connections were thus retained, corresponding to the amplitudes of the significant connections during that instant. A total of seven features were then calculated for each time point: the statistical moments up to order 4, the median, the maximum, and the range (the difference between the maximum and the minimum).

A support vector machine (Cortes and Vapnik 1995) with a radial basis function kernel was trained for each time point. We preferred a support vector machine (SVM) instead of a neural network in order to avoid overfitting, which is a typical problem of more complex models trained with a low number of trials. The training was carried out using a nested cross-validation procedure: (i) LOSO cross-validation was used to split the dataset into training (nine subjects) and test set (one subject), and (ii) for each fold of the LOSO cross-validation, ten fold cross-validation was used to furtherly divide the training set into training and validation set.

The inner validation loop was used to optimize the decoder hyperparameters and to perform feature selection through the minimum redundancy maximum relevance (Radovic *et al* 2017) algorithm.

The optimized hyperparameters were: C, i.e. the cost of misclassification of training instances; and the free parameter of the radial basis function gamma. Hyperparameter optimization of was carried out using a grid search on [0.001, 0.01, 0.1, 1, 10] for C and [0.001, 0.01, 0.1, 1] for gamma. For each fold of the outer validation loop (LOSO), the best hyperparameters were set as the C and gamma values which achieved the best mean accuracy in the inner ten-fold cross validation loop, thus resulting in a different

set of hyperparameters for each fold of the LOSO cross-validation.

For each fold, the accuracy was calculated on the test set. The time-varying accuracy was obtained by averaging the accuracies across all folds of the LOSO cross-validation procedure.

For each time point, the predicted labels were compared 1000 times with 1000 shuffled versions of the test set labels (NP or VP) to calculate the chance level. The procedure was repeated for each fold of the LOSO cross-validation, resulting in a null distribution of 1000 × (number of fold) accuracy values. An exact p-value was obtained by comparing the original accuracy with the null distribution.

The time-varying p-values were corrected for the multiple comparisons using a cluster-size-based statistical non-parametric mapping approach (Nichols and Holmes 2002) and deemed significant if lower than $\alpha = 0.05$.

Code implementation was based on the *scikit-learn* package for *Python* (Pedregosa *et al* 2011).

3.3. Quantification and statistical analysis

The non-normality of the data undergoing statistical testing was assessed using Shapiro–Wilk tests (Shapiro and Wilk 1965). Sizes *n1* and *n2* of the independent samples undergoing Mann–Whitney tests (Neuhäuser 2011) and the associated U statistics are reported in the results section as $U_{n1,n2} = U$. Statistical significance level α was 0.05. The inter-hemispheric significant connection that arose in one subject was not considered in the tests comparing connections in the DH versus connections in the NDH. Tests were computed using the *scipy* package for *Python* (Virtanen *et al* 2020).

4. Results

4.1. NPs and VPs elicit two unique networks

The neural networks elicited by the processing of NPs and VPs were investigated with SEEG. The data were recorded from ten Italian-native speaker patients with no language disorders who underwent surgical operation for drug-resistant epilepsy. NPs and VPs were encoded in the same acoustic stimulus and could be differentiated only by their syntactic context (some Italian homophonous phrases, such as *la porta* /la ˈpɔrta/—that can be interpreted either as a NP—'the door'—or a VP—'[s/he] brings her'). The complete list of stimuli is shown in table S1. After preprocessing, close recording contacts were arranged in groups called mini-regions of interest (mini-ROIs), each represented by a prototypical contact. The grouping resulted in a total of 396 mini-ROIs in the left—or DH and 577 mini-ROIs in the right—or NDH (figure 1(B)). To identify the networks involved in both NPs and VPs processing (i.e. the group of mini-ROIs bounded together by causal relations), we used PDC (Baccalá and Sameshima 2001) and a recently developed pipeline to determine the significance of ERC elicited by an RW (Cometa *et al* 2021).

We restricted the analysis to connections identified within the ultra-high gamma frequency band (150–300 Hz). In previous analysis (Artoni *et al* 2020), the signal recorded in the ultra-high gamma frequency band showed the greatest differentiation between NPs and VPs for most of the recording contacts. The pipeline discovered 13 significant connections for the NP case (2 in the DH and 11 in the NDH) and 20 connections for the VP condition (6 in the DH, 13 in the NDH, and 1 from the right temporal lobe to the left temporal lobe). We observed four connections active for both phrases in the NDH. Of these shared connections three were intra-temporal (figure 2(A)). Although the recording contacts were more in the NDH than in the DH (577 in the NDH and 396 in the DH), the ratio between the number of significant connections and the total number of channels was higher for the NDH ($4.16{\cdot}10^{-2}$ for the NDH and $2{\cdot}10^{-2}$ for the DH). The ratio between the number of channels participating in a significant connection and the total number of recording contacts in each lobe was the highest for the temporal lobes ($10.83{\cdot}10^{-2}$ for the NDH and $18.59{\cdot}10^{-2}$ for the NDH). For all the other lobes, this ratio was an order of magnitude lower. All the significant connections are shown in table S2.

We compared the estimated connections with the recorded CCEPs (Matsumoto and Kunieda 2019), which are an indicator of the presence of a direct cortico-cortical or cortico-subcortico-cortical anatomical pathway (Matsumoto *et al* 2004). We restricted the identification of the CCEPs only to pairs of channels forming significant connections. Out of 33 significant connections, 11 exhibited a CCEP. The contacts involved in a significant connection and with a relevant CCEP were placed closer together than those not showing CCEPs (Mann–Whitney $U_{22,11} = 53$, $p < 0.005$) (figure 2(B)).

Significant connections may be biased by clusters of closely placed contacts. Thus, to factor out a possible effect of this spatial sampling bias, we compared the distribution of the distances between pairs of contacts showing significant causal connections with the distribution of the distances between all channels (figure 2(C)). We did not detect any difference between the two distributions (Mann–Whitney $U_{29,47987} = 590819$, $p = 0.16$).

Finally, more significant connections in both NPs and VPs were found in subjects with electrodes placed in the NDH, in contrast to those with the DH implanted (Mann–Whitney $U_{4,5} = 18.5$, $p < 0.05$, figure 2(D)). This difference was still present even when normalizing the number of significant directed connections by the total amount of the possible connections for each subject (Mann–Whitney $U_{4,5} = 18$, $p < 0.05$). Only one subject had both hemispheres

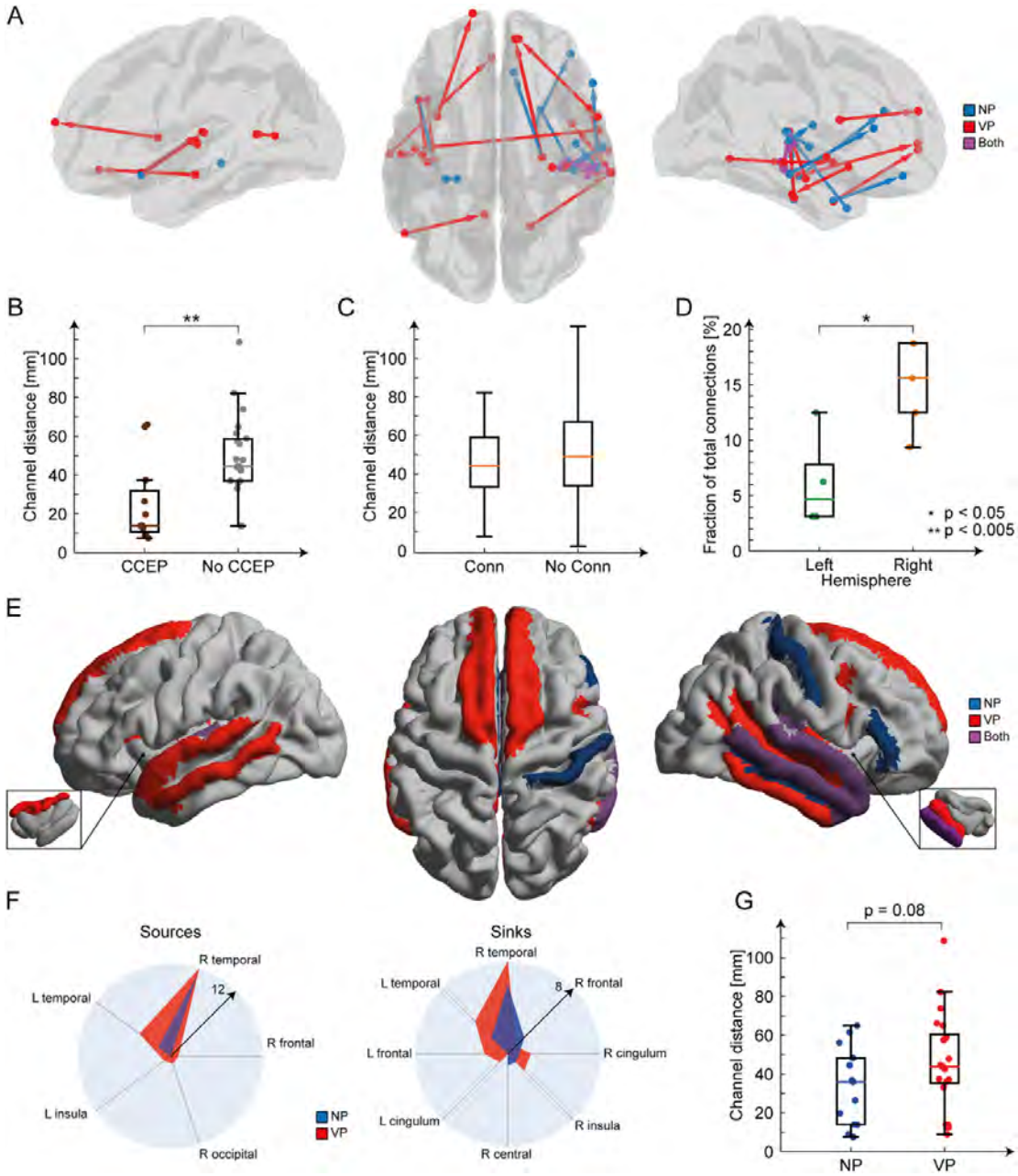

Figure 2. VPs engage a wider and more complex network. (A) Lateral and dorsal views of the identified directed connections. Nodes and edges are highlighted in blue for the noun phrase (NP) processing network, in red for the verb phrase related network, or purple if shared by both processing systems. (B) Box plots of the distances between the contacts involved in a significant connection and with a relevant cortico-cortical evoked potential (CCEP) and between those not showing CCEPs. (C) Box plots of the distances between pairs of implanted contacts, whether a significant connection exists between them (Conn) or not (No Conn). (D) Box plots of the number of connections in subjects with electrodes in the non-dominant (right) hemisphere and in those in which only the dominant (left) was probed. The vertical axis is normalized by the total number of significant directed connections identified across all subjects. (E) Lateral and dorsal views of the active brain zones during NPs (blue) processing, VPs (red) processing, or both (purple). An active brain zone is a cortical area containing one or more recording contacts that act as sources or sinks for a certain directed connection. The zoom-in pictures show the left and right insula. (F) Radar plots of the number of sources (left) and sinks (right) in each cerebral lobe, for the two conditions NP (blue) and VP (red). (G) Box plots of the distances between contacts involved in a significant connection during NP and VP processing.

implanted and showed an inter-hemispheric connection (VP, from the right temporal lobe to the left one).

4.2. VPs engage a wider network than NPs

The recording contacts participating in the NP-related network or the VP-related network were not spread across the entire cortical and subcortical volume but rather clustered in specific brain zones—i.e. the anatomical parcellation of cortical gyri and sulci according to the Destrieux atlas (Destrieux *et al* 2010). In total, 64 brain zones were probed in the DH and 88 in the NDH. Out of 152 cortical and subcortical areas, 11 were involved in the processing of both homophonous phrases (2 in the DH and 9 in the NDH), 12 participated in the processing of the VPs

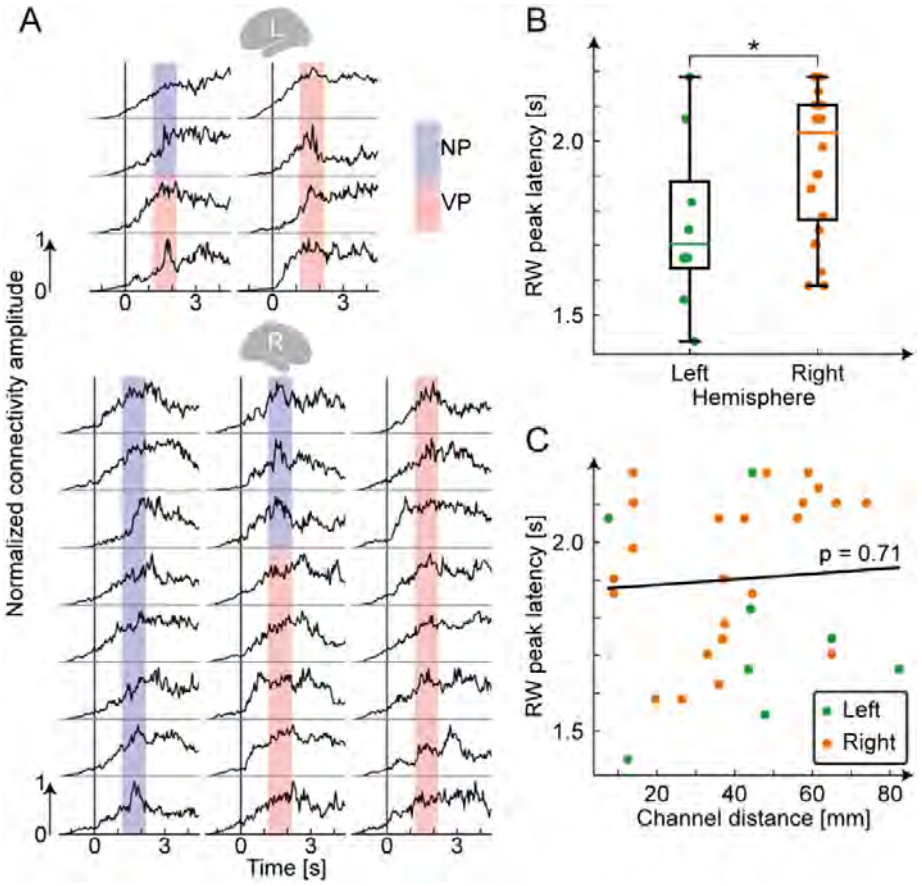

Figure 3. Syntax is faster in the dominant hemisphere. (A) Time series of the identified directed connections, in the dominant (L, top) and non-dominant (R, bottom) hemispheres. Each time series is normalized between 0 and 1. The 0 in the time axis is the start of the sentence, the colored area represents the homophonous part (or response window—RW). Directed connections that are significant when listening to noun phrases (NP) have this area coloured in blue, while those significant for verb phrases (VP) have this area highlighted in red. (B) Box plots of the latencies of the connectivity peaks during the RW in the dominant hemisphere (left) and in the non-dominant one (right). (C) Scatter plot of latencies of the peaks during the RW as function of the distances between channel pairs.

alone (6 in the DH and 6 in the NDH) and 6 responded exclusively to NPs (1 in the DH and 5 in the NDH) (figure 2(E)).

The connectivity estimated by the PDC is a directed causal information flow from one recording contact called source to another denoted sink. For NPs, all the sources were located bilaterally in the temporal lobes (2 in the DH and 11 in the NDH). For VPs, the temporal lobes contained 17 sources (5 in the DH and 12 in the NDH). The other three VPs sources were situated in the right occipital lobe, right frontal lobe, and left insula (figure 2(F), left). Most sinks, for both NPs and VPs, were in the two temporal lobes (DH: 2 for NPs and 4 for VPs; NDH: 6 for NPs and 8 for VPs). Other sinks were in the right insula (1 for NPs, 2 for VPs), in the right frontal lobe (2 for NPs, 1 for VPs), right central lobe (1 for NPs), right cingulum (1 for NPs, 2 for VPs), left frontal lobe (2 for VPs), and left cingulum (1 for VPs) (figure 2(F), right). The lists of the cortical and subcortical areas containing sources and sinks for a given connection are shown in tables S3 and S4.

Overall, VPs elicited more sources or sinks than NPs, engaged a higher number of different cortical and subcortical areas in both hemispheres, with almost no brain-zone being more active for NPs.

The results show that VPs extended the processing network beyond the temporal lobes.

Recording contacts that participated in VPs processing seemed to be located further than those involved in NPs processing (Mann–Whitney $U_{13,20} = 93$, $p = 0.08$, figure 2(G)), even if not reaching the statistical significance level $\alpha = 0.05$.

4.3. Syntax processing is faster in the DH

We then looked at the speed of response, or processing time, in the DH and NDH. The latencies of the peaks in the temporal evolutions of the time-varying significant causalities were thus compared among hemispheres. We smoothed the time-series with a Savgol filter in order to overcome the fluctuation of the neural signal and not altering the main peaks properties (Guiñón *et al* 2007, Benda and Volosyak 2019, Kawala-Sterniuk *et al* 2020). Then, we considered only the highest peak, for each smoothed time series, occurring during the homophonous part of the stimuli (figure 3(A)). These peaks arose earlier in the DH

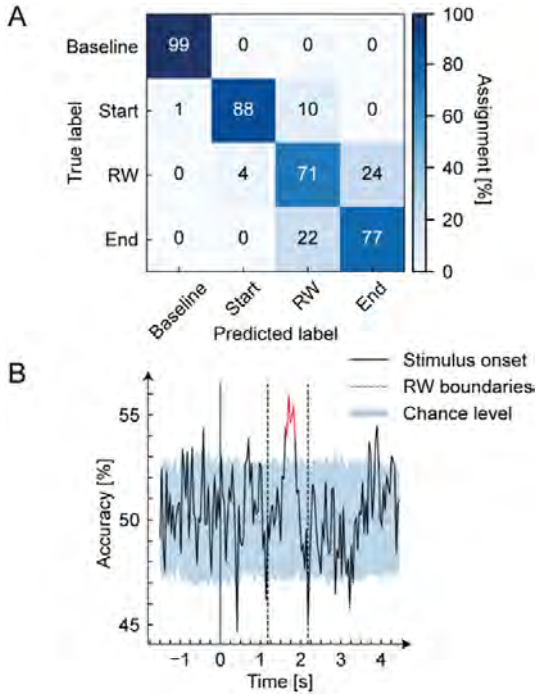

Figure 4. Connectivity decodes sentence structure. (A) Confusion matrix for the prediction of the stimulus phase. (B) Time-varying accuracy of the classification of NP vs VP. The blue line represents the median of the chance level, the boundaries of the light blue band are the 5th and 95th percentiles of the chance level distribution. The red part of the plot is the accuracy significantly above the chance level ($p < 0.005$).

(Mann–Whitney $U_{8,24} = 54.5, p < 0.05$), for both NPs and VPs (figure 3(B)).

The peak latencies in the directed connections evoked by the homophonous syntagms did not correlate linearly with the distances between the recording contacts involved in those connections (Pearson's $\rho = 0.07$, $p = 0.71$, figure 3(C)). Moreover, distances between recording contacts implanted in the DH and NDH and participating in an active connection were not statistically different (Mann–Whitney $U_{8,24} = 83$, $p = 0.29$). Therefore, the difference in peak latencies was likely not due to the channel distribution in the two hemispheres, but rather solely to the syntactic processing time.

4.4. Connectivity decodes homophonous phrases

We were interested in decoding the phase of the stimulus trial to test whether the time evolution of the PDC values carries information about the time evolution of the stimuli. The general neural connectivity estimated by the time-varying PDC was able to determine if the subject was waiting for the sentence (baseline), listening to the initial part of the sentence, to the homophonous phrase (RW), or its ending. We used a LSTM (Hochreiter and Schmidhuber 1997) to classify the stimulus segments with single-trial accuracy equal to 83.75% (the chance level is 38% due to class imbalance) (figure 4(A)).

We finally extracted time-dependent features only on the identified significant connections. We used a SVM (Cortes and Vapnik 1995) to predict the syntactic content of the homophonous phrase in the sentence. The accuracy was significantly above chance during the RW phase (figure 4(B)).

Both models were evaluated using a LOSO cross-validation.

5. Discussion

Language comprehension and production, in particularly syntax processing, are complex and highly integrated tasks continuously carried out by our brain, seemingly without effort. Analyzing their neural correlates thus requires sophisticated tools. One of the most promising techniques to identify the different neural processes underlying the syntactic operations leading to the processing of, for example, NP or VP is offered by directed connectivity evaluation related to the complexity of the large-scale networks. To our knowledge, this is the first time a difference in the connectivity elicited by NPs or VPs processing was identified.

Traditionally, the problem of understanding the neural correlates of syntax is approached by studying the effects of brain lesions or with syntax-related experimental tasks administered during neurophysiological and neuroimaging acquisitions contaminated by confounding factors such as phonology or semantics (Vigliocco *et al* 2011, Friederici *et al* 2017). Our approach is to leverage NP/VP homophonous phrases. The advantage of our solution is that we can factor out phonological and morphological confounding factors by analyzing these homophonous phrases.

The shift from the analysis of isolated lexical elements such as bare Vs and Ns vs. syntactic units, namely VPs and NPs, is obviously a necessary step toward the goal of capturing syntactic information. Lexical elements in isolation contain linguistic information but these pieces of information are artificially expressed in single words whereas natural linguistic expressions always involve syntactic computation. In fact, the stimuli involved syntax in two directions: first, each homophonous phrase was syntactically connected with other words expressing a full-fledged sentence; second, each homophonous phrase contained very different syntactic structures. More specifically: in NPs the surfacing order of the two words composing them, namely an article and a noun, was the same as the underling structure composing it; in VPs, the situation is completely different and definitely more complex. In all VPs considered here a transformation called *cliticization* takes place. The order of the elements constituting it (a pronoun, playing the role of the object, and a verb) is reversed with respect to the canonical order in an SVO language like Italian; the canonical position of the object

is to the right of the verb (Moro 2016). All in all, the shift from V/N to VP/NP constitutes a necessary and relevant step towards the final goal of cracking the underlying code of human syntax.

5.1. Decoding of the syntactic category and potential applications for BCIs

The information carried by all the directed connections was able to discriminate between parts of the sentence. The syntactic category of the stimulus was discriminable just by looking at the significant connections, showing that restricting the topology analysis on the few significant connections allows to decode NPs vs. VPs while keeping a lower computational complexity. The computational complexity is one of the key factors that should be controlled in the development of an online speech decoder.

In recent years, there have been important technological and methodological advancements in perceived and imagined speech decoding (Martin *et al* 2018, Panachakel and Ramakrishnan 2021). Recent works focus on the classification of vowels (Duc and Lee 2020, Mahmud *et al* 2020), syllables (Brandmeyer *et al* 2013, Correia *et al* 2015, Archila-Meléndez *et al* 2018), words (Ossmy *et al* 2015, Vorontsova *et al* 2021, Proix *et al* 2022) and complete sentences (Zhang *et al* 2012, Chakrabarti *et al* 2015), distinguishing stimuli mainly at the semantic level. The most advanced online decoding techniques rely heavily on the articulatory representation of syllables and words in the motor and supplementary motor cortices (Anumanchipalli *et al* 2019, Moses *et al* 2021). However, this approach can only be applied to patients with intact motor commands, but who are unable to move the muscles that are necessary for speech production, which represent a minority of the patients with speech impairment (Guenther *et al* 2009, Wilson *et al* 2020). Thus, other decoding strategies that rely on the brain regions that encode speech are needed (Proix *et al* 2022).

Here, we decoded the syntactic category of the homophonous part of the acoustic stimuli exploiting 29 different speech-encoding cortical and subcortical areas spanning the entire brain. Only recently such strategy has been used in the decoding of groups of syllables and words (Proix *et al* 2022).

However, our approach relies on the time evolution of the connectivity values between recording contacts. This solution has the advantage of assuring high inter-subject generalizability as shown by the LOSO validation results: the connectivity features are independent of the location of the implanted leads, which may differ from subject to subject. Also, our method is well suited to be implemented in an online decoder. Moreover, the signals that drive the decoding are directly entangled to the syntactic representation of the stimuli rather than their phonological—and articular—components.

We believe that a decoding strategy that relies on multiple language-encoding cortical and subcortical areas will drastically improve the performance of speech prostheses and may be the key missing piece for the development of this technology. There are, however, a number of limitations that will need to be addressed in the future to fully exploit this strategy. The most critical among them are: the computational complexity needed to calculate the causality between a large number of recording contacts, the need to cover wide parts of the brain, even if SEEG represents a very promising technique due to its relatively low invasiveness (Cometa *et al* 2022).

5.2. Describing two syntax-related neural networks

We identified a low number of significant connections compared to all the possible ones. This is not surprising, since the human cortex seems to be sparsely connected (Rosen and Halgren 2022).

We showed that VPs processing, compared to NPs processing, elicited a significantly higher number of directed connections, linked together more brain structures both in the DH and in the NDH, and involved the activation of a wider cortical and subcortical network. VPs processing was distributed beyond temporal lobes, pushing the information from sources located in the right frontal lobe and left insula, to sinks in both frontal lobes, anterior cingulate regions, and right insula. This suggests a greater network small-worldness for NPs, with a preference for short-range connections over long-range ones.

Most of the literature converges on a more extended cerebral involvement in verb processing than for nouns (Vigliocco *et al* 2011, Lukic *et al* 2021). However, again, most evidence came from tasks requiring the processing of N/V as words in isolation: this is the first time an approach based on homophonous phrases, hence syntax, is used.

Temporal lobes (both in the DH and in the NDH) seem to be the main hub in which the syntactic operations leading to NPs or VPs are analyzed and processed. For NPs all the information flow started from these areas, while for VPs 3 out of 20 sources were placed outside the temporal lobes (with the one in the right occipital cortex very close to temporal areas). Also, sinks were mostly located in the temporal lobes. The important role of the temporal lobes, in particular of left posterior regions, in syntactic processing is supported by lesion and imaging evidence (Friederici *et al* 2017, Matchin and Hickok 2020).

The comparison of the estimated directed connections with the CCEPs arising between recording contacts showed a partial discrepancy. While the structural connectivity underlying CCEPs is well known (e.g. the Human Connectome Project) (Van Essen *et al* 2012), the functional and

effective connectivity are patterns of highly heterogeneous causal relationships that may reflect processes occurring during many different temporal time scales (Vincent *et al* 2007, Shmuel and Leopold 2008, Honey *et al* 2009, Matsui *et al* 2011, Keller *et al* 2014). The ERC identified here, is thus the expression of more complex neural processes, for which there are no unique *a priori* hypotheses. However, measures based on the Granger causality framework such as the PDC used here were shown to describe well the interactions occurring between coupled neural populations (Kamiński *et al* 2001, Cadotte *et al* 2008).

Interestingly, recording contacts involved in a significant connection and showing at the same time CCEPs were implanted closed together than the pairs of channels without relevant CCEPs. Indeed, CCEPs may terminate their propagation early (Logothetis *et al* 2010, Keller *et al* 2014), which is in agreement with the description of CCEPs as supported by short-range local relations arising from direct hard-wired connections via cortico-cortical or cortico-subcortico-cortical pathways (Matsumoto *et al* 2004). This suggests that syntax-related processing relies mostly on long-range connections between cortical or subcortical areas, expressing network-level neural synchronization supported by long-range, indirect structural pathways, typical of high-level cognitive processing (Salmelin and Kujala 2006).

We attempted to counteract the imbalance of implanted electrodes in the different lobes and hemispheres through the use of mini-ROIs and by applying statistical tests on normalized measures. However, with magnetic resonance imaging (MRI) recordings of all subjects, a spatial modeling of the sampled neural activity could have been used to handle this issue (Esposito *et al* 2013, Singer *et al* 2014).

5.3. The role of the two hemispheres

Earlier peaks in the connectivity time-series in the DH revealed that the syntax processing elicited by our stimuli started first in the temporal lobes of the left hemispheres and then spread to the right cortices. The directed links from DH to NDH that are necessary to transfer the information from one hemisphere to the other were not deemed significant because they were probably active during all sentence processing, and so they were masked during the search for the causal connections with the highest amplitude increase during the homophonous part of the stimulus. Also, only one subject out of ten was implanted in both hemispheres.

Focal lesion, behavioral, fMRI and electrophysiological studies provide converging evidence for a dominant role of one hemisphere (the left in right-handers and in the majority of left-handers) for most aspects of language processing (Tzourio-Mazoyer *et al* 2017). Here we detected more significant connections arising in the NDH than in the DH. Focal lesion, behavioral, fMRI and electrophysiological studies provide converging evidence for a dominant role of one hemisphere (the left in right-handers and in the majority of left-handers) for most aspects of language processing (for a recent review, see Tzourio-Mazoyer *et al* 2017). While speech perception is often considered as a bi-hemispheric process (Hickok and Poeppel 2000, Poeppel *et al* 2008; see however Scott and McGettigan 2013), syntactic processing is strongly associated with left hemispheric function (Matchin and Hickok 2020, Grodzinsky *et al* 2021). Our finding of more significant connections arising in the NDH than in the DH is thus unexpected. Additional work is needed to better characterize the role of the NDH in syntax processing.

5.4. Surprisal and other confounding factors

It has not escaped our attention the fact that our results concerning syntactic structures converge with parsing as shown by a surprisal analysis (Artoni *et al* 2020). Syntactic surprisal is related to the expectedness of a given word's syntactic category given its preceding context and it is based on the frequency of the occurrence within a Corpus. These models however are limited and cannot fully capture syntactic dependencies as they involve hierarchical relations such as those expressed in phrases. In our previous paper (Artoni *et al* 2020) we showed that surprisal values could be sorted into VPs and NPs classes at best by means of Support Vector Machine Analysis with a score of only 86%; also while significant differences were found considering the surprisal of the articles and the clitics no significant difference in surprisal could be seen between the verbs and nouns, indicating that surprisal alone cannot fully explain VPs and NPs differences. The unresolved tension between syntax and surprisal deserves at least three important remarks: first, surprisal is a measure of the probability for a word to occur after another in a given corpus collecting real expressions of a language whereas modern linguistics aims at understanding what is not produced in a given language as generated by a given grammar; second, Markovian chains models, upon which this kind of surprisal is formalized, have been proved not to be able to capture the structure of natural languages (ever since the pioneering work of Noam Chomsky in the late fifties); third, there are indeed other models of surprisal involving hierarchical relations such as those expressed in phrases but they obviously rely on syntactic structures, such as the relation between a head and a complement yielding a phrase, and this does not show that surprisal is sufficient to understand linguistic regularities: rather it shows that for these regularities to be captured, syntactic notions must be exploited. Recently,

deep neural networks were used to model the surprisal (Goldstein *et al* 2022, Heilbron *et al* 2022, Russo *et al* 2022), showing their ability to explain the neural activity elicited by sentence processing. These neural networks merge different linguistic information to estimate the surprisal: it would be surprising if the hierarchical relations and thus the syntactic information would not have been exploited for this estimation.

This crucial issue has not been solved yet in the current debate and we can only expect our work to contribute to this by offering hints for a future clarification, among other things.

In light of these considerations, although we did not include any non-syntactic condition in our experiment, we consider unlikely that the response we observe is not syntactic-specific. This is because, although it is clear that nouns which refer to objects (say, table) are semantically poorer even than a relatively simple verb (say, *destroy*) in that they completely lack theta-roles such as agent, patient, etc, it is also true that there are nouns which do come with the same richness in terms of theta-roles (say, *destruction*). The relative simplicity of nouns over the verbs is still to be understood and this preliminary work is also to be expanded in that direction, aiming at a comprehensive distinction between nouns vs. verbs. Moreover, although it is surely true that verbs comes with a more complex paradigm with respect to nouns, it is not true that the a V is necessarily more complex than a verb from a morphological point of view. All words exploited here, practically, consist of a lexical morpheme and a functional one. For example, for nouns: the lexical root *port-* as in *porta* (door) with the singular—feminine morpheme *-a*; for verbs, the lexical root *port-* as in *porta* (s/he brings) and the third singular present *-a*. Strictly speaking, then the morphological differences were reduced to double morpheme constructions. All in all, the fact that the homophonous elements are morphologically comparable suggests that the contrast between verbs and nouns activation is arguably devoted to the operation of cliticization involving reordering of words with verbs vs. basic order with nouns. In fact, it would be very surprising that such a complex operation as cliticization would not require a higher activation.

Furthermore, the lexical or semantic differences present in the stimuli do not systematically reflect any of the dimensions with a known impact on brain activity, such as length, syllabic structure, frequency, familiarity, semantic category, imageability, valence, arousal, etc.

6. Conclusions

In our previous work (Artoni *et al* 2020) we identified the high-gamma activity as the main neural correlate of syntactic processing. However, we failed at characterizing the network involved in syntactic processing. Treating the recording sites and thus the corresponding cortical hubs as segregated structures cannot truly describe the neural processes responsible for the syntactic representation of NPs and VPs in the human brain. Here we expanded the previous work by considering the set of causal connections arising between cortical and sub-cortical structures during syntactic processing. This method allowed us not only to identify the sites in which such processing occurs but also to describe how these sites communicate between them. For example, we show a preference for a posterior-to-anterior pathway for the neural connections, mainly from the temporal lobes to the frontal lobes. Knowing how brain structures communicate when performing a cognitive task may help clinicians and engineers developing treatments and technologies for language impairment mitigation. We show the plausibility of a decoding strategy that relies on the temporal evolution of the causal connections calculated between recording channels. This decoding strategy does not depend on the subject-specific recording sites, thus allowing a simplification of the calibration procedure of neuro-prostheses for language impairment mitigation. Furthermore, we compared the connections identified using the PDC with those arising from electrical stimulation, i.e. the CCEP. This comparison allowed us to hypothesize the anatomical features supporting the syntax-related neural networks, i.e. indirect long-range structural pathways.

In conclusion, these results give an unprecedented overview of the mechanisms involved in the neural representation of the syntactic structures as they represent an important step forward in human language comprehension, contributing to the full characterization of syntactic processing. We showed a specific brain activity encoding a syntactic distinction, which is faster in the DH. Since, even from a purely formal point of view, syntactic processing cannot be compared with other computational systems, language-related or not (Chomsky 2014, Moro 2014a, 2014b), it is reasonable to conclude that the network highlighted here is not only specific but arguably it is uniquely dedicated to syntax. We prove that it is possible to decode the syntactic structure of a phrase by looking at the connections elicited by speech processing between multiple cortical and subcortical areas. This could contribute to the future development of speech prostheses for speech impairment mitigation (Anumanchipalli *et al* 2019).

Data availability statement

The data cannot be made publicly available upon publication due to legal restrictions preventing unrestricted public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

The work was financed by:

- 'Inner Speech Converter Technology' (INSPECT) PRIN 2017JPMW4F project funded by the Italian Ministry of University and Research.
- 'La neurolinguistica federata, dalle strutture al senso figurato' PRO3 project funded by the Italian Ministry of University and Research.
- 'A multiscale integrated approach to the study of the nervous system in healt and disease' (MNESYS) and 'Tuscany Health Ecosystem' (THE) projects funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5 by the Italian Ministry of University and Research and European Union—NextGenerationEU.
- 'Bertarelli foundation'.

Author contributions

Conceptualization: S F C, A M, and S M; Methodology: A C, F A, F B, C R, G L R, S F C, A M, and S M; Experimental setup: F A, Data Collection: F A, Software: A C and F A; Validation: A C, F A, P d O, and M R; Formal Analysis: A C; Theoretical linguistics contribution: A M, Investigation: P d O and M R; Resources: G L R and S M; Data Curation: A C, F A, P d O, and M R; Writing - Original draft: A C, F A; Writing - Review and Editing: A C, F A, S F C, A M, and S M; Visualization: A C; Supervision: G L R, S F C, A M, and S M; Project Administration: S F C, A M, and S M; Funding Acquisition: S M, F A.

ORCID iDs

Andrea Cometa https://orcid.org/0000-0002-5771-0316
Fiorenzo Artoni https://orcid.org/0000-0002-0967-6643

References

Abadi M *et al* 2015 *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems* (available at: www.tensorflow.org/)
Anumanchipalli G K, Chartier J and Chang E F 2019 Speech synthesis from neural decoding of spoken sentences *Nature* **568** 493–8
Archila-Meléndez M E, Valente G, Correia J M, Rouhl R P W, van Kranen-mastenbroek V H and Jansma B M 2018 Sensorimotor representation of speech perception. cross-decoding of place of articulation features during selective attention to syllables in 7T fMRI *Eneuro* **5** 2
Artoni F *et al* 2020 High gamma response tracks different syntactic structures in homophonous phrases *Sci. Rep.* **10** 7537
Artoni F, Fanciullacci C, Bertolucci F, Panarese A, Makeig S, Micera S and Chisari C 2017 Unidirectional brain to muscle connectivity reveals motor cortex control of leg muscles during stereotyped walking *NeuroImage* **159** 403–16
Astolfi L *et al* 2008 Tracking the time-varying cortical connectivity patterns by adaptive multivariate estimators *IEEE Trans. Biomed. Eng.* **55** 902–13
Baccalá L A and Sameshima K 2001 Partial directed coherence: a new concept in neural structure determination *Biol. Cybern.* **84** 463–74
Batterink L and Neville H J 2013 The human brain processes syntax in the absence of conscious awareness *J. Neurosci.* **33** 8528–33
Benda M and Volosyak I 2019 Peak detection with online electroencephalography (EEG) artifact removal for brain–computer interface (BCI) purposes *Brain Sci.* **9** 347
Brandmeyer A, Farquhar J D R, McQueen J M and Desain P W M 2013 Decoding speech perception by native and non-native speakers using single-trial electrophysiological data *PLoS One* **8** e68261
Brunner M I, Bárdossy A and Furrer R 2019 Technical note: stochastic simulation of streamflow time series using phase randomization *Hydrol. Earth Syst. Sci.* **23** 3175–87
Buzsáki G and Draguhn A 2004 Neuronal oscillations in cortical networks *Science* **304** 1926–9
Cadotte A J, DeMarse T B, He P and Ding M 2008 Causal measures of structure and plasticity in simulated and living neural networks *PLoS One* **3** e3355
Chakrabarti S, Sandberg H M, Brumberg J S and Krusienski D J 2015 Progress in speech decoding from the electrocorticogram *Biomed. Eng. Lett.* **5** 10–21
Chomsky N 2014 *Aspects of the Theory of Syntax* vol 11 (Cambridge, MA: MIT Press)
Cometa A, D'Orio P, Revay M, Micera S and Artoni F 2021 Stimulus evoked causality estimation in stereo-EEG *J. Neural Eng.* **18** 056041
Cometa A, Falasconi A, Biasizzo M, Carpaneto J, Horn A, Mazzoni A and Micera S 2022 Clinical neuroscience and neurotechnology: an amazing symbiosis *iScience* **25** 105124
Correia J M, Jansma B M B and Bonte M 2015 Decoding articulatory features from fmri responses in dorsal speech regions *J. Neurosci.* **35** 15015
Cortes C and Vapnik V 1995 Support-vector networks *Mach. Learn.* **20** 273–97
Cossu M *et al* 2015 Stereoelectroencephalography-guided radiofrequency thermocoagulation in the epileptogenic zone: a retrospective study on 89 cases *J. Neurosurg.* **123** 1358–67
Cossu M, Cardinale F, Castana L, Citterio A, Francione S, Tassi L, Benabid A L and Lo Russo G 2005 Stereoelectroencephalography in the presurgical evaluation of focal epilepsy: a retrospective analysis of 215 procedures *Neurosurgery* **57** 706–18
Crowther L J, Brunner P, Kapeller C, Guger C, Kamada K, Bunch M E, Frawley B K, Lynch T M, Ritaccio A L and Schalk G 2019 A quantitative method for evaluating cortical responses to electrical stimulation *J. Neurosci. Methods* **311** 67–75
Destrieux C, Fischl B, Dale A and Halgren E 2010 Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature *NeuroImage* **53** 1–15
Ding N, Melloni L, Zhang H, Tian X and Poeppel D 2015 Cortical tracking of hierarchical linguistic structures in connected speech *Nat. Neurosci.* **19** 158–64
Do T-T N, Lin C-T and Gramann K 2021 Human brain dynamics in active spatial navigation *Sci. Rep.* **11** 13036
Duc N T and Lee B 2020 Decoding brain dynamics in speech perception based on eeg microstates decomposed by multivariate Gaussian hidden Markova model *IEEE Access* **8** 146770–84
Engel A K and Singer W 2001 Temporal binding and the neural correlates of sensory awareness *Trends Cogn. Sci.* **5** 16–25
Esposito F, Singer N, Podlipsky I, Fried I, Hendler T and Goebel R 2013 Cortex-based inter-subject analysis of iEEG and fMRI data sets: application to sustained task-related BOLD and gamma responses *NeuroImage* **66** 457–68
Fedorov A *et al* 2012 3D slicer as an image computing platform for the quantitative imaging network *Magn. Reson. Imaging* **30** 1323–41

Fischl B 2012 FreeSurfer *NeuroImage* **62** 774–81
Friederici A D 2018 The neural basis for human syntax: Broca's area and beyond *Curr. Opin. Behav. Sci.* **21** 88–92
Friederici A D, Chomsky N, Berwick R C, Moro A and Bolhuis J J 2017 Language, mind and brain *Nat. Hum. Behav.* **1** 713–22
Friederici A D and Kotz S A 2003 The brain basis of syntactic processes: functional imaging and lesion studies *NeuroImage* **20** S8–17
Friederici A D, Wang Y, Herrmann C S, Maess B and Oertel U 2000 Localization of early syntactic processes in frontal and temporal cortical areas: a magnetoencephalographic study *Hum. Brain Mapp.* **11** 1–11
Geweke J 1982 Measurement of linear dependence and feedback between multiple time series *J. Am. Stat. Assoc.* **77** 304–13
Goldstein A *et al* 2022 Shared computational principles for language processing in humans and deep language models *Nat. Neurosci.* **25** 369–80
Granger C W J 1969 Investigating causal relations by econometric models and cross-spectral methods *Econometrica* **37** 424
Griffiths J D, Marslen-Wilson W D, Stamatakis E A and Tyler L K 2013 Functional organization of the neural language system: dorsal and ventral pathways are critical for syntax *Cereb. Cortex* **23** 139–47
Grodzinsky Y and Friederici A D 2006 Neuroimaging of syntax and syntactic processing *Curr. Opin. Neurobiol.* **16** 240–6
Grodzinsky Y, Pieperhoff P and Thompson C 2021 Stable brain loci for the processing of complex syntax: a review of the current neuroimaging evidence *Cortex* **142** 252–71
Guenther F H *et al* 2009 A wireless brain-machine interface for real-time speech synthesis *PLoS One* **4** e8218
Guiñón J, Ortega E, García-Antón J and Pérez-Herranz V 2007 Moving average and Savitzki-Golay smoothing filters using Mathcad *ICEE 2007* (*Coimbra, Portugal, 3–7 September 2007*) vol 2007 pp 1–4 (available at: http://icee2007.dei.uc.pt/)
Gwin J T, Gramann K, Makeig S and Ferris D P 2011 Electrocortical activity is coupled to gait cycle phase during treadmill walking *NeuroImage* **54** 1289–96
He B, Astolfi L, Valdes-Sosa P A, Marinazzo D, Palva S O, Benar C G, Michel C M and Koenig T 2019 Electrophysiological brain connectivity: theory and implementation *IEEE Trans. Biomed. Eng.* **66** 2115–37
Heilbron M, Armeni K, Schoffelen J-M, Hagoort P and de Lange F P 2022 A hierarchy of linguistic predictions during natural language comprehension *Proc. Natl Acad. Sci.* **119** e2201968119
Hickok G and Poeppel D 2000 Towards a functional neuroanatomy of speech perception *Trends Cogn. Sci.* **4** 131–8
Ho Y and Wookey S 2020 The real-world-weight cross-entropy loss function: modeling the costs of mislabeling *IEEE Access* **8** 4806–13
Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80
Honey C J, Sporns O, Cammoun L, Gigandet X, Thiran J P, Meuli R and Hagmann P 2009 Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl Acad. Sci. USA* **106** 2035–40
Kamiński M, Ding M, Truccolo W A and Bressler S L 2001 Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance *Biol. Cybern.* **85** 145–57
Kawala-Sterniuk A, Podpora M, Pelc M, Blaszczyszyn M, Gorzelanczyk E J, Martinek R and Ozana S 2020 Comparison of smoothing filters in analysis of EEG data for the medical diagnostics purposes *Sensors* **20** 807
Kayne R S 2019 What is suppletion? On *goed and on went in modern English *Trans. Phil. Soc.* **117** 434–54
Keller C J, Honey C J, Mégevand P, Entz L, Ulbert I and Mehta A D 2014 Mapping human brain networks withcortico-ortical evoked potentials *Phil. Trans. R. Soc.* B **369** 20130528
Knösche T R, Maess B and Friederici A D 1999 Processing of syntactic information monitored by brain surface current density mapping based on MEG *Brain Topogr.* **12** 75–87
Korzeniewska A, Crainiceanu C M, Kuś R, Franaszczuk P J and Crone N E 2008 Dynamics of event-related causality in brain electrical activity *Hum. Brain Mapp.* **29** 1170–92
Lachaux J P, Rudrauf D and Kahane P 2003 Intracranial EEG and human brain mapping *J. Physiol.* **97** 613–28
Logothetis N K, Augath M, Murayama Y, Rauch A, Sultan F, Goense J, Oeltermann A and Merkle H 2010 The effects of electrical microstimulation on cortical signal propagation *Nat. Neurosci.* **13** 1283–91
Lukic S *et al* 2021 Dissociating nouns and verbs in temporal and perisylvian networks: evidence from neurodegenerative diseases *Cortex* **142** 47–61
Magrassi L, Aromataris G, Cabrini A, Annovazzi-Lodi V and Moro A 2015 Sound representation in higher language areas during language generation *Proc. Natl Acad. Sci.* **112** 1868–73
Mahmud M S, Ahmed F, Yeasin M and Bidelman G M 2020 Decoding categorical speech perception from evoked brain responses *2020 IEEE Region 10 Symp. (TENSYMP)* pp 766–9
Maris E and Oostenveld R 2007 Nonparametric statistical testing of EEG- and MEG-data *J. Neurosci. Methods* **164** 177–90
Martin S, Iturrate I, Millán J D R, Knight R T and Pasley B N 2018 Decoding inner speech using electrocorticography: progress and challenges toward a speech prosthesis *Front. Neurosci.* **12** 422
Matchin W and Hickok G 2020 The cortical organization of syntax *Cereb. Cortex* **30** 1481–98
Matsui T, Tamura K, Koyano K W, Takeuchi D, Adachi Y, Osada T and Miyashita Y 2011 Direct comparison of spontaneous functional connectivity and effective connectivity measured by intracortical microstimulation: an fMRI study in macaque monkeys *Cereb. Cortex* **21** 2348–56
Matsumoto R, Kunieda T and Nair D 2017 Single pulse electrical stimulation to probe functional and pathological connectivity in epilepsy *Seizure* **44** 27–36
Matsumoto R and Kunieda T 2019 Cortico-cortical evoked potentials mapping *Invasive Studies of the Human Epileptic Brain—Principles and Practice* ed S Lhatoo, P Kahane and H Lüders (Oxford : Oxford University Press) pp 431–52
Matsumoto R, Nair D R, LaPresto E, Najm I, Bingaman W, Shibasaki H and Lüders H O 2004 Functional connectivity in the human language system: a cortico-cortical evoked potential study *Brain* **127** 2316–30
Milde T, Leistritz L, Astolfi L, Miltner W H R, Weiss T, Babiloni F and Witte H 2010 A new Kalman filter approach for the estimation of high-dimensional time-variant multivariate AR models and its application in analysis of laser-evoked brain potentials *NeuroImage* **50** 960–9
Moro A 2014a On the similarity between syntax and actions *Trends Cogn. Sci.* **18** 109–10
Moro A 2014b Response to pulvermüller: the syntax of actions and other metaphors *Trends Cogn. Sci.* **18** 221
Moro A 2016 *Impossible Languages* (Cambridge, MA: MIT Press) (https://doi.org/10.7551/mitpress/9780262034890.001.0001)
Moses D A *et al* 2021 Neuroprosthesis for decoding speech in a paralyzed person with Anarthria *New Engl. J. Med.* **385** 217–27
Munari C, Hoffmann D, Francione S, Kahane P, Tassi L, Lo Russo G and Benabid A L 1994 Stereo-electroencephalography methodology: advantages and limits *Acta Neurol. Scand.* **152** 56–67

Narizzano M, Arnulfo G, Ricci S, Toselli B, Tisdall M, Canessa A, Fato M M and Cardinale F 2017 SEEG assistant: a 3DSlicer extension to support epilepsy surgery *BMC Bioinform.* **18** 124

Neuhäuser M 2011 Wilcoxon–Mann–Whitney test *International Encyclopedia of Statistical Science* ed M Lovric (Berlin: Springer) pp 1656–8

Nichols T E and Holmes A P 2002 Nonparametric permutation tests for functional neuroimaging: a primer with examples *Hum. Brain Mapp.* **15** 1–25

Nordin A D, Hairston W D and Ferris D P 2019 Human electrocortical dynamics while stepping over obstacles *Sci. Rep.* **9** 4693

Nuñez S C, Dapretto M, Katzir T, Starr A, Bramen J, Kan E, Bookheimer S and Sowell E R 2011 fMRI of syntactic processing in typically developing children: structural correlates in the inferior frontal gyrus *Dev. Cogn. Neurosci.* **1** 313–23

Ossmy O, Fried I and Mukamel R 2015 Decoding speech perception from single cell activity in humans *NeuroImage* **117** 151–9

Panachakel J T and Ramakrishnan A G 2021 Decoding covert speech from EEG-A comprehensive review *Front. Neurosci.* **15** 392

Pedregosa F *et al* 2011 Scikit-learn: machine learning in python *J. Mach. Learn. Res.* **12** 2825–30

Penny W D, Stephan K E, Mechelli A and Friston K J 2004 Modelling functional integration: a comparison of structural equation and dynamic causal models *NeuroImage* **23** S264–74

Poeppel D, Idsardi W J and van Wassenhove V 2008 Speech perception at the interface of neurobiology and linguistics *Phil. Trans. R. Soc.* B **363** 1071–86

Proix T *et al* 2022 Imagined speech can be decoded from low- and cross-frequency intracranial EEG features *Nat. Commun.* **13** 48

Pylkkänen L 2019 The neural basis of combinatory syntax and semantics *Science* **366** 62–66

Radovic M, Ghalwash M, Filipovic N and Obradovic Z 2017 Minimum redundancy maximum relevance feature selection approach for temporal gene expression data *BMC Bioinform.* **18** 9

Rosen B Q and Halgren E 2022 An estimation of the absolute number of axons indicates that human cortical areas are sparsely connected *PLoS Biol.* **20** e3001575

Russo A G, Ciarlo A, Ponticorvo S, Di Salle F, Tedeschi G and Esposito F 2022 Explaining neural activity in human listeners with deep learning via natural language processing of narrative text *Sci. Rep.* **12** 17838

Russo S *et al* 2021 Focal lesions induce large-scale percolation of sleep-like intracerebral activity in awake humans *NeuroImage* **234** 117964

Salmelin R and Kujala J 2006 Neural representation of language: activation versus long-range connectivity *Trends Cogn. Sci.* **10** 519–25

Schell M, Zaccarella E and Friederici A D 2017 Differential cortical contribution of syntax and semantics: an fMRI study on two-word phrasal processing *Cortex* **96** 105–20

Schwarz G 1978 Estimating the dimension of a model *Ann. Stat.* **6** 461–4

Scott S K and McGettigan C 2013 Do temporal processes underlie left hemisphere dominance in speech perception? *Brain Lang.* **127** 36–45

Shapiro S S and Wilk M B 1965 An analysis of variance test for normality (complete samples) *Biometrika* **52** 591–611

Shmuel A and Leopold D A 2008 Neuronal correlates of spontaneous fluctuations in fMRI signals in monkey visual cortex: implications for functional connectivity at rest *Hum. Brain Mapp.* **29** 751–61

Singer N, Podlipsky I, Esposito F, Okon-Singer H, Andelman F, Kipervasser S, Neufeld M Y, Goebel R, Fried I and Hendler T 2014 Distinct iEEG activity patterns in temporal-limbic and prefrontal sites induced by emotional intentionality *Cortex* **60** 121–38

Trebaul L *et al* 2018 Probabilistic functional tractography of the human cortex revisited *NeuroImage* **181** 414–29

Tzourio-Mazoyer N, Perrone-Bertolotti M, Jobard G, Mazoyer B and Baciu M 2017 Multi-factorial modulation of hemispheric specialization and plasticity for language in healthy and pathological conditions: a review *Cortex* **86** 314–39

Van Essen D C *et al* 2012 WU-Minn HCP Consortium The human connectome project: a data acquisition perspective *NeuroImage* **62** 2222–31

Varela F, Lachaux J P, Rodriguez E and Martinerie J 2001 The brainweb: phase synchronization and large-scale integration *Nat. Rev. Neurosci.* **2** 229–39

Vigliocco G, Vinson D P, Druks J, Barber H and Cappa S F 2011 Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies *Neurosci. Biobehav. Rev.* **35** 407–26

Vincent J L, Patel G H, Fox M D, Snyder A Z, Baker J T, Van Essen D C, Zempel J M, Snyder L H, Corbetta M and Raichle M E 2007 Intrinsic functional architecture in the anaesthetized monkey brain *Nature* **447** 83–86

Virtanen P *et al* 2020 SciPy 1.0: fundamental algorithms for scientific computing in Python *Nat. Methods* **17** 261–72

Vorontsova D *et al* 2021 Silent EEG-speech recognition using convolutional and recurrent neural network with 85% accuracy of 9 words classification *Sensors* **21** 6744

Weber K, Christiansen M H, Petersson K M, Indefrey P and Hagoort P 2016 FMRI syntactic and lexical repetition effects reveal the initial stages of learning a new language *J. Neurosci.* **36** 6872

Wilson G H, Stavisky S D, Willett F R, Avansino D T, Kelemen J N, Hochberg L R, Henderson J M, Druckmann S and Shenoy K V 2020 Decoding spoken English from intracortical electrode arrays in dorsal precentral gyrus *J. Neural Eng.* **17** 066007

Wu J, Ngo G H, Greve D, Li J, He T, Fischl B, Eickhoff S B and Yeo B T T 2018 Accurate nonlinear mapping between MNI volumetric and FreeSurfer surface coordinate systems *Hum. Brain Mapp.* **39** 3793–808

Xia M, Wang J and He Y 2013 BrainNet viewer: a network visualization tool for human brain connectomics *PLoS One* **8** e68910

Zhang D, Gong E, Wu W, Lin J, Zhou W and Hong B 2012 Spoken sentences decoding based on intracranial high gamma response using dynamic time warping *Proc. Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society, EMBS* pp 3292–5