Decoding semantics from dynamic brain activation patterns: From trials to task in EEG/MEG source space

Federica Magnabosco[1], Olaf Hauk[1]

[1] MRC Cognition and Brain Sciences Unit, University of Cambridge, UK

Key words:

semantic cognition, MVPA, decoding, task, semantics, visual word recognition

Abstract

The temporal dynamics within the semantic brain network and its dependence on stimulus and task parameters are still not well understood. Here, we addressed this by decoding task as well as stimulus information from source-estimated EEG/MEG data. We presented the same visual word stimuli in a lexical decision (LD) and three semantic decision (SD) tasks. The meanings of the presented words varied across five semantic categories. Source space decoding was applied over time in five ROIs in the left hemisphere (Anterior and Posterior Temporal Lobe, Inferior Frontal Gyrus, Primary Visual Areas, and Angular Gyrus) and one in the right hemisphere (Anterior Temporal Lobe). Task decoding produced sustained significant effects in all ROIs from 50-100 ms, both when categorising tasks with different semantic demands (LD-SD) as well as for similar semantic tasks (SD-SD). In contrast, semantic word category could only be decoded in lATL, rATL, PTC and IFG, between 250-500 ms. Furthermore, we compared two approaches to source space decoding: Conventional ROI-by-ROI decoding and combined-ROI decoding with back-projected activation patterns. The former produced more reliable results for word-category decoding while the latter was more informative for task-decoding. This indicates that task effects are distributed across the whole semantic network while stimulus effects are more focal. Our results demonstrate that the semantic network is widely distributed but that bilateral anterior temporal lobes together with control regions are particularly relevant for the processing of semantic information.

Significance Statement

Most previous decoding analyses of EEG/MEG data have focussed on decoding performance over time in sensor space. Here for the first time we compared two approaches to source space decoding in order to reveal the spatio-temporal dynamics of both task and stimulus features in the semantic brain network. This revealed that even semantic tasks with similar task demands can be decoded across the network from early latencies, despite reliable differences in their evoked responses. Furthermore, stimulus features can be decoded in both tasks but only for a subset of ROIs and following the earliest task effects. These results inform current neuroscientific models of controlled semantic cognition.

Introduction

Semantic cognition is the representation and processing of the acquired knowledge about the world. The semantic brain network that enables us to store, employ, manipulate, and generalise conceptual knowledge can be characterised along two dimensions: representation and control (Lambon Ralph et al., 2017). Bilateral anterior temporal lobes (ATLs) constitute a multimodal hub within the representation component, where modality-specific information is combined to represent coherent concepts (Patterson et al., 2007). However, context and task influence the type of information that is relevant in any particular moment, requiring a control system capable of manipulating and shaping the activations in the representation system (Jefferies, 2013). A recent meta-analysis indicated that the control system is mostly left-hemispheric and comprises inferior frontal gyrus (IFG) and posterior temporal cortex (PTC) (Jackson, 2021). However, little is known about the dynamics within and between the components of the semantic network. In this study, we are tapping into semantic brain dynamics by decoding task as well as stimulus features from source-estimated EEG/MEG data in a visual word recognition paradigm.

In neuroimaging research, EEG/MEG and fMRI datasets have traditionally been analysed with different decoding/multi-voxel-pattern-analysis approaches (Grootswagers et al., 2017). While fMRI studies often focus on *spatial information* of decoding accuracies across different cortical regions (e.g., Haxby et al., 2001; Huth et al., 2016), EEG/MEG studies have mostly focussed on the *time cours*e of decoding accuracy determined in sensor space (e.g., Cichy et al., 2014). However, ROI-wise decoding in source space is possible using appropriate distributed source estimation methods (e.g., Kietzmann et al., 2019). Alternatively, it has been proposed to interpret the topography of classifier weights from the decoding analysis, and e.g. submit them to source estimation. In this case, the weights have to be back-projected into activation patterns before the application of source estimation (Haufe et al., 2014). This back-projection method should be applied whenever the classifier patterns themselves (rather than decoding accuracy) are interpreted, since otherwise seemingly high contributions from some sensors or voxel may be due to noise rather than signal (Haufe et al., 2014). This offers another approach to decoding analysis in source space: One can apply decoding to data across all ROIs that are assumed to be involved in the processes of interest, and interpret the distribution of these decoding weights across ROIs following back-projection, thus determining which regions contribute more or less to the successful decoding of stimulus or task information. To our knowledge, these two approaches – conventional ROI-by-ROI decoding and across-ROI decoding with back projection – have not been applied to the same EEG/MEG dataset in source space yet.

The anterior temporal lobes are sensitive to the semantic characteristics of single words as shown in several EEG/MEG studies both in univariate analyses (Dhond et al., 2007; Farahibozorg et al., 2022; Marinkovic et al., 2003) as well as in multivariate analyses on intracranial recordings (Chan et al., 2011; Rogers et al., 2021).

A recent study examined the activation dynamics in several brain areas within the semantic network by contrasting two different tasks that differed with respect to the depth of semantic processing: semantic decision (SD) and lexical decision (LD) (Rahimi et al., 2022). The evoked responses indicated that different semantic task demands are associated with an early modulation of visual and attentional processes, followed by differences in the semantic information retrieval in the ATLs, and finally a modulation in control regions (PTC and IFG) involved in the extraction of task-relevant features for response selection. These results were corroborated by the functional connectivity analysis in this and another study that revealed connectivity between the ATLs and semantic control regions (Rahimi et al., 2022; Rahimi, Jackson, Farahibozorg, et al., 2023).

The present study aimed to extend the evoked results of Rahimi et al. (2022) using a multivariate decoding approach in source space. In contrast to the previous study, we investigated both stimulus and task features. In addition, we evaluated two different decoding approaches: per-ROI and across-ROI. We used these methods to compare the spatio-temporal decodability of stimulus and task features of dynamic brain activation patterns. In particular, we asked 1) whether the patterns that carry stimulus and task information overlap in space and time, and to what extent this information is distributed vs. localised; and 2) whether task affects the amount of available stimulus-specific information within the semantic network.

Methods

We used the dataset described in Farahibozorg (2018) and Rahimi et al. (2022), and most pre-processing steps are identical to that study. Thus, we will here report a summary of the most relevant characteristics and refer to the previous study for more detailed information. Previous analyses of this dataset did not employ decoding of task or stimulus features. The code for the analysis is available on https://github.com/magna-fede/SourceSpaceDecoding_SDvsLD.

Participants

We analysed data from 18 participants. All of them were native English speakers, right-handed, with normal or corrected-to-normal vision, and reported no history of neurological

disorders or dyslexia. The study was approved by the Cambridge Psychology Research Ethics Committee.

Stimuli and Procedure

250 uninflected words were used in the visual stimulation. Each word belonged to one of 5 different categories based on its semantic content: visual, auditory, hand-actions, emotional or neutral abstract words. Table 1 summarizes the psycholinguistic variables. For the stimulus feature decoding, we used only a subset of the stimuli. Specifically, we created a subset of the two abstract words categories (i.e., emotional and neutral categories), so that we ended up with 4 semantic categories with 50 stimuli each. This was done because the stimuli in the emotional and neutral categories contained 1.5 letters more on average, compared to the other three categories. To avoid a false semantic classification due to word length, we removed this confound by selecting 50 words across these two categories so to have a better match for possible confounds, resulting in a more generic abstract word category.

The study consisted of four blocks and the full set of word stimuli was presented in each of them. One block consisted of a lexical decision (LD) task, where participants were presented with 250 words and an additional 250 pseudowords. They were asked to press a button with either their left index or ring finger to indicate whether the letter string represented a real, meaningful word or not. The other three blocks were all different instances of a semantic decision (SD) task: Participants were presented with the same 250 words as for LD, plus 50 additional fillers and 30 target words. Participants had to press a button with their left middle finger whenever the current stimulus belonged to the specific target semantic category for that block. The three target semantic categories were: (1) “food that contains milk, flour, or egg”, (2) “non-citrus fruits”, and (3) “something edible with a distinctive odour”. The target categories were unrelated to the word categories of the stimuli described above. The order of presentation of the SD blocks was randomised; half of the participants performed the LD task first, and half of them did it after the SD blocks.

In the SD block, only non-target trials were included in the analyses, i.e. trials that did not require a button press response. In the LD task participants were responding with a button press to both real words and pseudowords since a two-alternative forced choice is the standard procedure in LD tasks. We do not consider the details of response execution at the end of each trial as a serious confound for our EEG/MEG results in earlier latency ranges (see Rahimi et al., 2022 for details). These potential confounds are present only when classifying task in SD vs LD blocks, but not when considering the classification of different SD blocks or when performing single-word semantic classification.

	Number of letters	Frequency	Orthographic Neighbourhood Size	Bigram Frequency	Trigram Frequency
Action	5.02 (1.20)	19.67 (30.86)	5.44 (5.25)	17368.95 (9580.53)	1690.77 (2505.16)
Auditory	5.16 (1.22)	11.03 (17.25)	5.28 (5.24)	17599.13 (9181.36)	1825.17 (2426.59)
Visual	5.18 (1.48)	13.32 (17.58)	4.50 (5.44)	20089.52 (11698.66)	1867.21 (3314.68)
Abstract	5.28 (0.93)	23.64 (23.76)	3.26 (3.84)	19495.50 (9564.39)	1597.65 (955.62)
Pseudowords	5.00 (1.00)	N/A	4.70 (4.50)	19465.83 (10435.78)	1670.66 (2029.80)

Table 1. Summary statistics (mean and (standard deviation)) for each of the (single-word) semantic categories presented in both the SD and LD task, and for pseudowords presented only in the LD task.

Data acquisition and source estimation

Simultaneous EEG and MEG data were concurrently recorded to maximize spatial resolution (Molins et al., 2008). The sampling rate during data acquisition was 1000 Hz and an online bandpass filter 0.03 to 330 Hz was applied. Signal Space Separation with its spatiotemporal extension as implemented in the Neuromag Maxwell-Filter software was applied to the raw MEG data. Raw data were visually inspected for each participant, and bad EEG channels were marked and linearly interpolated. Data were then band-pass filtered between 0.1 and 45 Hz. FastICA algorithm was applied to the filtered data to remove eye movement and heartbeat artefacts. We used L2-Minimum Norm Estimation (MNE) for source reconstruction (Hämäläinen & Ilmoniemi, 1994; Hauk, 2004). Three-layer boundary element forward models were constructed using structural MRI scans.

Regions of interest

We focused our analyses on six regions of interest that were defined using the anatomical masks provided by the Human Connectome Project (HCP) parcellation (Glasser et al., 2016). We adhered to Rahimi's et al. (2022) selection, which examined the evoked responses and functional connectivity analysis of these same ROIs (Figure 1). These areas were chosen due to their putative role in the semantic brain network, as derived from neuroimaging fMRI studies. ATLs have been proposed to be multimodal semantic hub regions (Lambon Ralph et al., 2017). IFG (Inferior Frontal Gyrus) and PTC (Posterior

Temporal Cortex) have been described as semantic control regions required for the appropriate task- or context-relevant processing of a word stimulus (Jackson, 2021). Angular Gyrus (AG) has been suggested as an additional hub region or convergence zone (Binder, 2016). Finally, we included PVA (Primary Visual Areas) to test potential task effects on early perceptual processing.

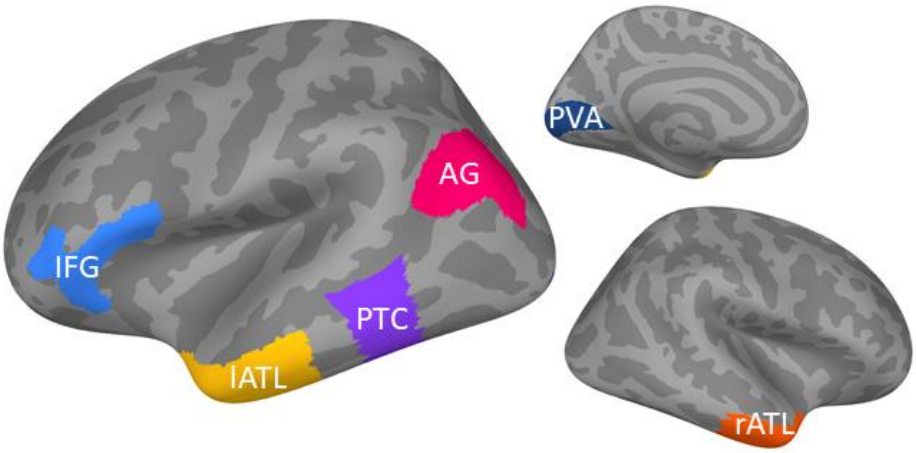


Figure 1. Regions of interest (ROIs) used in this study (based on Rahimi et al. (2022)).

Pre-processing

The pre-processing and decoding analyses are based on tutorials and examples available on MNE-Python (Gramfort et al., 2013) and on the tutorial provided by Grootswagers et al. (2017). We estimated brain responses in source space for each participant and for each trial. We sampled our data in epochs that started from 300 ms before the stimulus onset until 900 ms post-stimulus. Data was down-sampled to 250 Hz. For the decoding analysis, we intended to keep as much information in our activation patterns as possible and therefore used signed source estimates, i.e. rather than taking the (only positive-valued) intensities of dipole sources at each vertex, we kept their directions of current flow. We present the grand-averaged time courses of brain activation for each ROI and for different tasks in Figure 2 (using the “mean-flip” option to account for the variability of source orientations). The figures show that we obtained clear evoked responses in all ROIs and conditions.

The pre-processing consisted in averaging 3 trials together for each class, in order to reduce noise in the data and improve decoding accuracy (Isik et al., 2014). We then z-score-normalised each of the resulting epochs. The models fitted were always logistic regression models, using a five-fold cross-validation procedure. We calculated accuracy as the average Receiver Operator Characteristic (ROC) Area-Under-Curve (AUC) on the test data across the 5 folds.

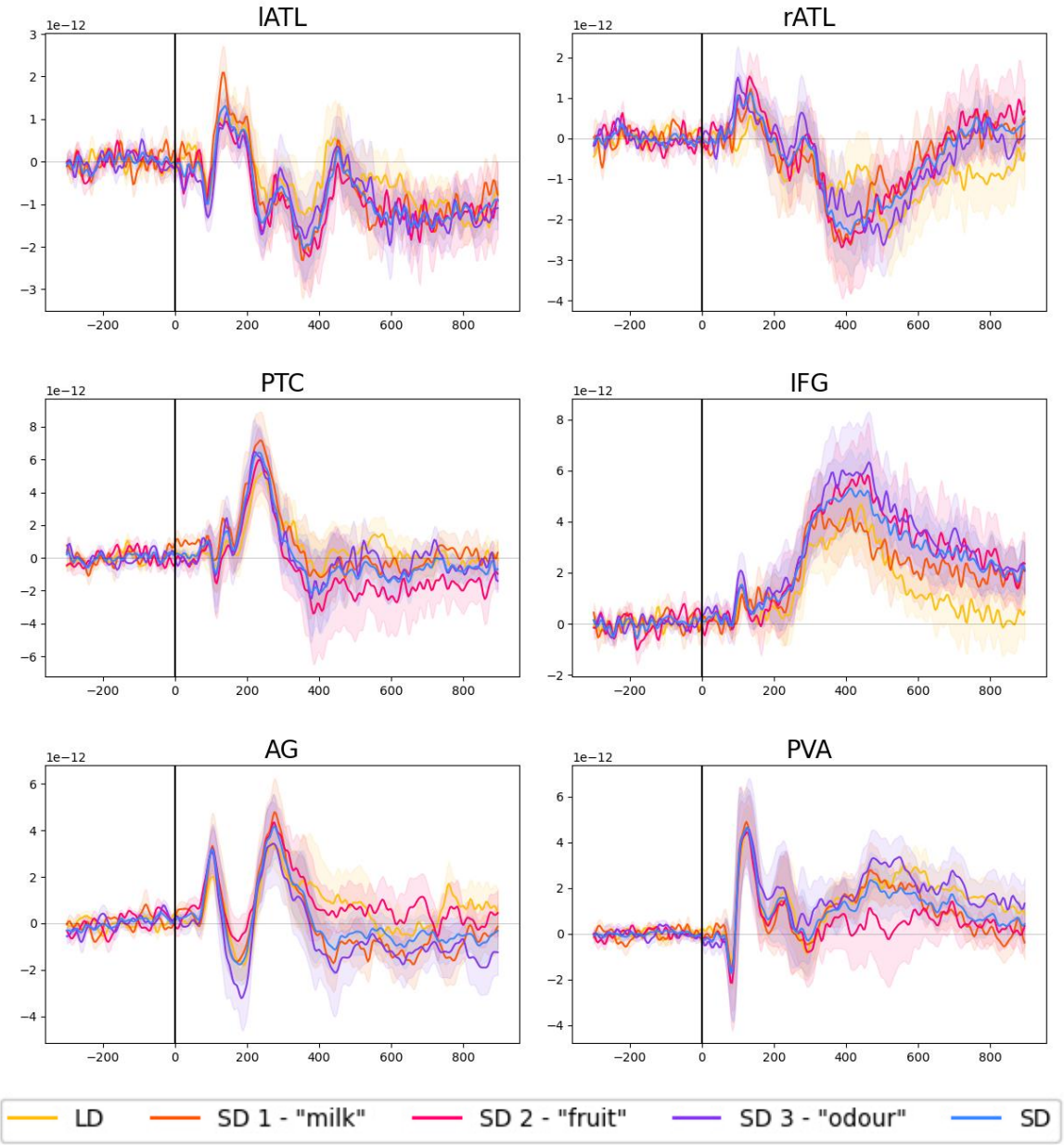


Figure 2. Evoked responses for each task and semantic decision tasks averaged together. Shadowed area report standard error of the mean (LD: lexical decision; SD: semantic decision; "milk", "fruit" and "odour" decision refer to single SD blocks).

Decoding Analysis

We first decoded task in two different ways: LD-SD, where we classified whether a trial belonged to a lexical decision or semantic decision block; and SD-SD, where we decoded to which of the 3 SD blocks a certain trial belonged. Then, we decoded single-word semantic category, separately for LD and SD blocks. For each classification analysis, we compared two different decoding approaches: A) Individual ROIs: we estimated decoding accuracy in each ROI separately; B) Combined ROIs: we fitted a model that used all vertices across all

ROIs, then determined the contribution of each ROI by looking at the root-mean-squared of the back-projected weights of the model within ROIs.

Individual ROIs Accuracy

In the LD-SD classification, we separately classified each of the LD-SD combinations using a binary logistic regression model, and accuracy was calculated as the average across them. In the SD-SD and stimulus feature classification we fitted multinomial logistic regression models and calculated accuracies as one-vs-rest ROC AUC (this ensured that the chance level was 0.5 for all tasks). For the semantic category analysis, we fitted a multiclass logistic regression model. For SD blocks, we estimated accuracy both when concatenating the 3 blocks in one model (in order to have more trials, and therefore a more reliable estimate of regions and time points where stimulus features were decodable) as well as separately per block (to have a fair comparison in terms of noise level with the LD block with a comparable number of trials).

We performed statistical significance analyses to test differences in accuracy between different classification tasks, specifically, we tested A) whether task decoding accuracy differed between LD-SD and SD-SD; B) whether we could reliably decode word category classification above chance; C) whether stimulus feature decoding accuracy differed between LD and SD tasks. We used a cluster-based permutation test over time to correct for multiple comparisons (Maris & Oostenveld, 2007).

Combined ROIs Patterns

In this analysis, we decoded patterns combined across ROIs and then determined how much each ROI contributes to the corresponding classifier. For this purpose, we back-projected the classifier weights into the corresponding activation patterns across ROIs as suggested by Haufe et al. (2014). This back-projection is required because the classifier weights may reflect both signal and noise, while the back-projected patterns reflect brain activity predicted by the estimated model.

In general, we followed the same steps as in the individual ROIs procedure described in the previous section, and subsequently extracted the back-projected weights of each feature in the model, which we will refer to as activation patterns or just patterns from now on.

For each classification task, we will report the root-mean-squared (RMS) activation patterns for each ROI, averaged across participants. As each model had a set of weights specific to each class, we will first calculate an ROI's RMS of each class-specific activation pattern – for example in the semantic classification, ROI-specific hand-class patterns (and analogously for visual, auditory, and abstract classes). Then, by averaging across classes, we obtain the

average activation pattern time course of each ROI, for each participant. The RMS pattern is a measure of the relative contribution of a certain ROI to the classification: Each vertex's value reflects the relative contribution of that specific vertex to the signal used by the classifier, i.e. vertices that are close to 0 are relatively less informative, whereas larger absolute values indicate that the vertex contains more information. By considering the RMS instead of the average, we avoid that patterns of vertices within the same ROI with different polarities will cancel each other.

Results

In the following, we will present results for different task classifications (LD-SD and SD-SD; Figure 3) and semantic word category classification (Figure 4). For each part, we will first present the results for individual-ROI classification followed by combined-ROI classification.

Task classification

Figure 3 shows classification accuracy results for the task comparisons (left: LD-SD, right: SD-SD) over time for different ROIs in source space, separately for the individual-ROI approach (A) and the combined-ROI approach (B and C).
Both task classifications show similar results, in so far as task was decodable from before 100ms after stimulus presentation and then for most of the epoch, both in the individual- and combined-ROI approach. Detailed information is presented below.

Individual ROIs Accuracy

In LD-SD, accuracy is at chance level in the baseline period (before stimulus presentation), but it sharply increases after approximately 50 ms post-stimulus reaching accuracies of over 0.7 in all examined ROIs at around 100 ms. Accuracy plateaus for a large part of the epoch and is followed by a slow decline. The SD-SD classification shows very similar results, although accuracy is generally lower, peaking below 0.7. Cluster-based permutation testing (not shown) confirms this difference with LD-SD accuracy being significantly higher than SD-SD in all ROIs from around 70 ms post-stimulus presentation to the end of the epoch (apart from IFG, where the difference starts at approximately 152 ms).

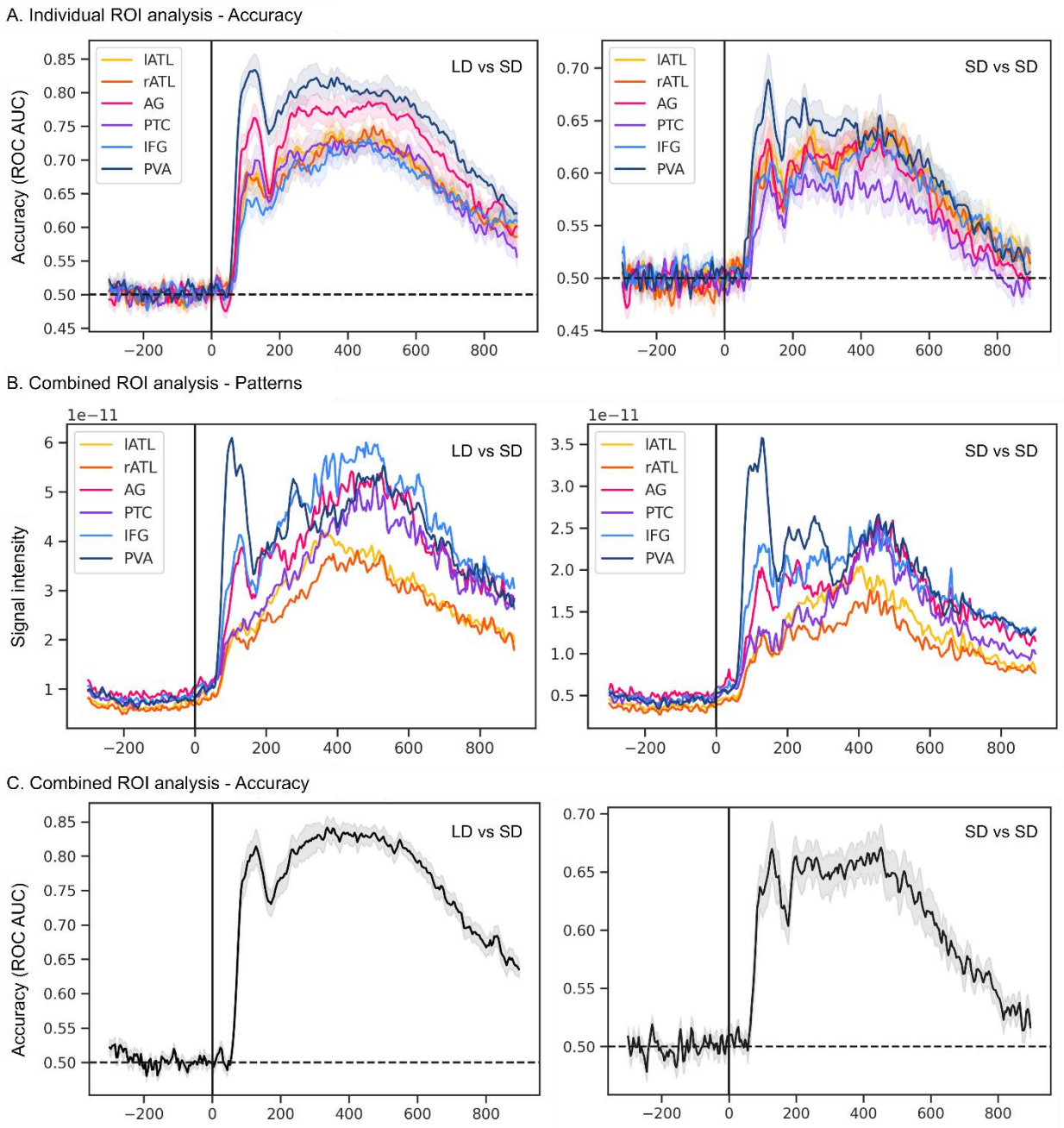


Figure 3. Task classification performance. For all plots, the shadowed area represents the standard error of the mean (across participants). ***A.*** *Individual ROI task classification accuracy (ROC AUC) when decoding LD from SD (left) and when decoding SD tasks (right).* ***B.*** *Root-mean-square of the activation patterns within each ROI, when decoding LD from SD (left) and when decoding SD tasks (right).* ***C.*** *Combined ROI task classification accuracy (ROC AUC) when decoding LD from SD (left) and when decoding SD tasks (right).*

Combined ROIs Patterns

When combining all ROIs in one model, the accuracy profile is very similar to that in the individual-ROI approach, i.e. it is characterized by a sharp increase soon after stimulus onset and a plateau afterwards (Figure 3C). However, while in the previous analysis the

shape of accuracy time courses was very similar across ROIs, we observed larger differences among ROIs when examining the RMS of activation patterns within ROIs (Figure 3B). For example, PVA's (dark blue line) contribution to task classification (both LD-SD and SD-SD) peaks around 100ms. AG's and IFG's time courses (pink and light blue, respectively) are similar, with a smaller peak than PVA around 100ms, but similar and even larger values after 250 ms. In contrast, the temporal lobes (lATL and rATL, yellow and orange) and posterior temporal cortex (PTC, purple) show similar time courses without a distinct early peak. Their activations slowly increase until they reach their peak around 400 ms. Also in this case, cluster-based permutation testing (not shown) showed that decoding accuracy was higher in LD-SD and SD-SD from approximately 64 ms post-stimulus onset. Thus, while the Individual-ROI analyses revealed that all ROIs carry information about the tasks, the Combined-ROI provides more information about the time courses of information, in particular about their differential contributions at early (~100ms) versus later (>250ms) latencies.

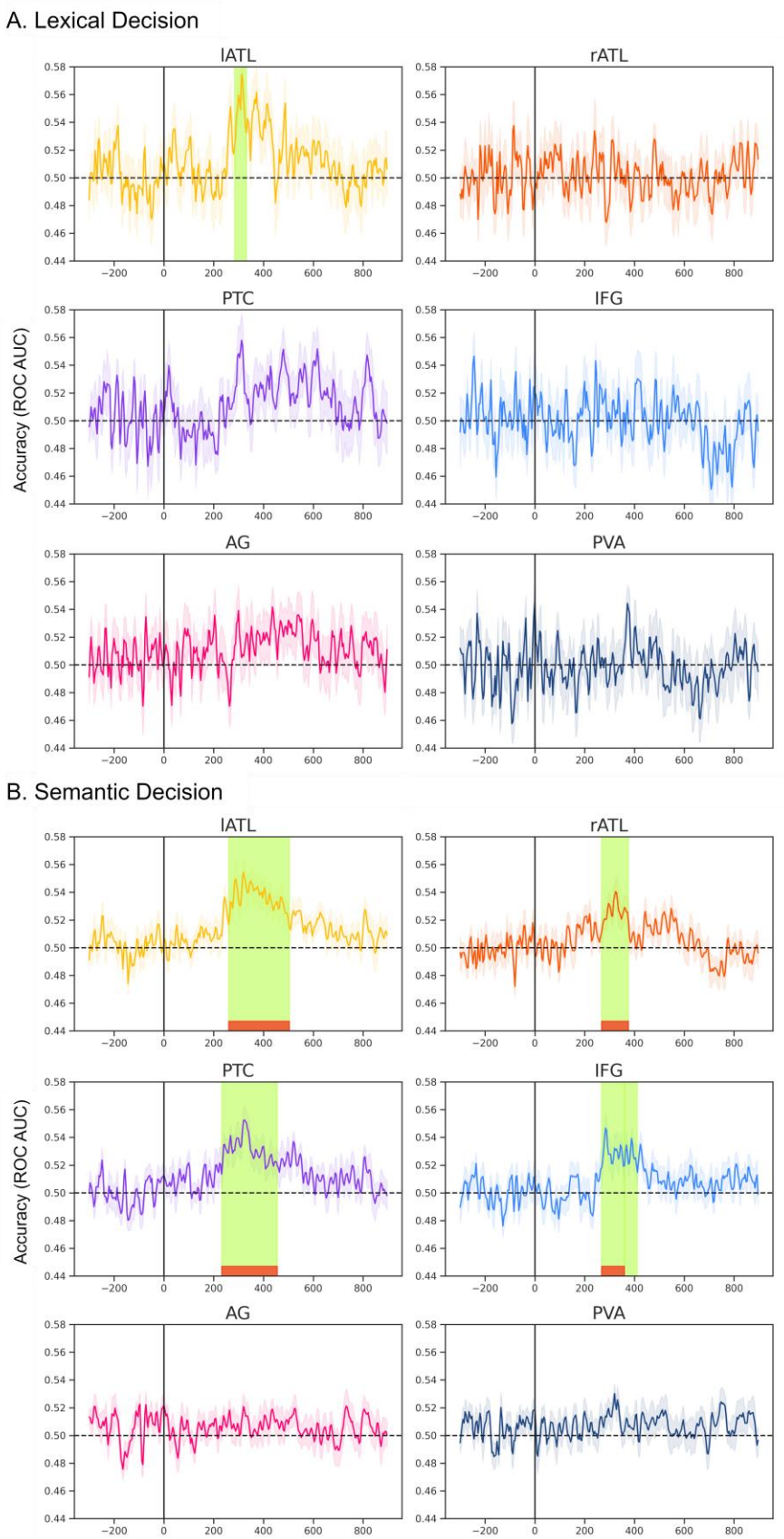
Single-word semantic features classification

In the semantic category classification, for each stimulus we decoded the semantic category it belonged to (out of 4 alternatives), separately for LD and SD blocks. Figures 4A and 4B show results for the individual ROIs stimulus feature decoding. Figures 4C and 4D show results for the combined ROIs pattern and accuracy. In general, we observed that stimulus features were decodable above chance level when participants were engaged in a semantic decision task, but less so in lexical decision.

Individual ROIs Accuracy

Figure 4A revealed that in LD blocks only lATL showed significant decoding for word category and only in one short time window around 300 ms. In contrast, for SD we obtained significant decoding in several regions (left and right ATL, PTC and IFG). SD decoding survived Bonferroni correction across ROIs. For each ROI we report the significant cluster(s) and their approximate latencies in Table 2. In SD, we observed above-chance decoding accuracy in left and right ATL, PTC and IFG, but not in PVA and AG. PTC successfully classifies single-word semantic category between 236-452 ms post-stimulus. This is followed by left and right ATL, respectively, between 264-500 and 272-372 ms. IFG was above chance approximately at the same time, between 282-356 and 364-408 ms (but the second cluster did not survive Bonferroni correction). As the SD model was trained on trials from 3 blocks, the table reports also the results of cluster-based permutation tests for each SD block separately, to show results from a comparable noise level to the single LD block.

Decoding accuracy in single SD blocks was above chance in more ROIs-latency combinations compared to the LD block (apart from the odour task).



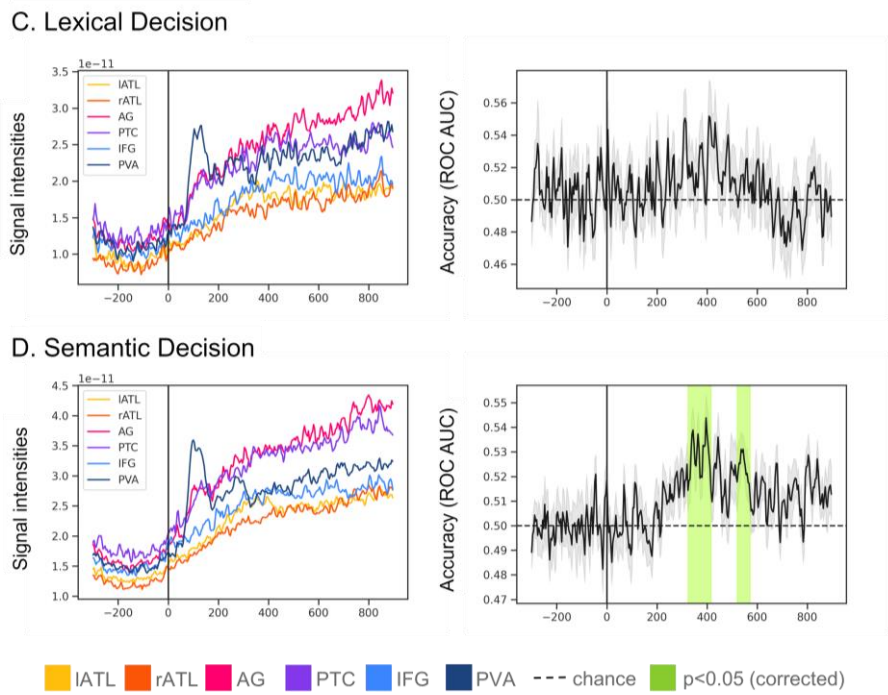


Figure 4. Single-word semantic category classification performance. For all plots, the shadowed area represents the standard error of the mean (across participants). Green colour highlights the times when cluster-based permutation correction test revealed a significant cluster (α=0.05) where accuracy was above 0.5, and red the clusters that are still significant after Bonferroni correction (across ROIs). ***A.*** *Individual ROI task classification accuracy (ROC AUC) when decoding semantic features in LD block.* ***B.*** *Individual ROI task classification accuracy (ROC AUC) when decoding semantic features in SD blocks.* ***C.*** *Combined ROI root-mean-square of the activation patterns within each ROI (left) and classification accuracy (ROC AUC) (right), when decoding semantic features in LD block when decoding LD.* ***D.*** *Combined ROI root-mean-square of the activation patterns within each ROI (left) and classification accuracy (ROC AUC) (right), when decoding semantic features in SD block.*

	lATL	rATL	AG	PTC	IFG	PVA
LD	288-328 ms	-	-	-	-	-
SD	264-500 ms	272-372 ms	-	236-452 ms	272-356 ms 364-408 ms	-
Bonferroni Corrected	264-500 ms	272-372 ms	-	236-452 ms	272-356 ms	-
SD 1 - "milk"	-	316-372 ms	716-760 ms	-	588-632 ms	-
SD 2 - "fruit"	348-404 ms	296-360 ms 544-592 ms	428-476 ms	316-408 ms 440-492 ms 788-844 ms	-	320-376 ms
SD 3 - "odour"	-	-	-	-	248-292 ms	-

Table 2. Approximate timings when single-word semantic category is decodable in each ROI. SD refers to the information when fitting a model to all SD blocks concatenated, whereas the effects for models fitted to individual SD blocks (i.e., (1) "milk", (2) "fruit", and (3) "odour") are reported in the last three rows of the table.

We then statistically tested whether the decoding accuracy was significantly higher in SD than in LD considering that word category was decodable only in lATL in the LD block, in contrast to all the semantic network in SD. However, no ROI showed a significant effect. To further understand whether task affected the stimulus category decoding, we computed cross-task decodability. This analysis was included to determine whether the representation of specific semantic information was consistent across tasks (i.e., successful cross-task decoding indicates that multivariate patterns of a certain word category are represented similarly irrespective of tasks). We trained a model on SD trials and tested stimulus features decodability of the same model in LD trials, and vice versa. Results are reported in Figure 5. When the model was trained on SD and tested on LD trials, we observed above-chance accuracy in all semantic areas: lATL between 276-408 ms, rATL between 240-456 ms, PTC between 272-512 ms, IFG between 360-408 ms, and PVA between 184-240 ms (where the last two regions did not survive Bonferroni correction across ROIs). We observed similar results in models trained on LD and tested on SD (although LD trials were significantly less so we will not interpret the results for each individual ROI).

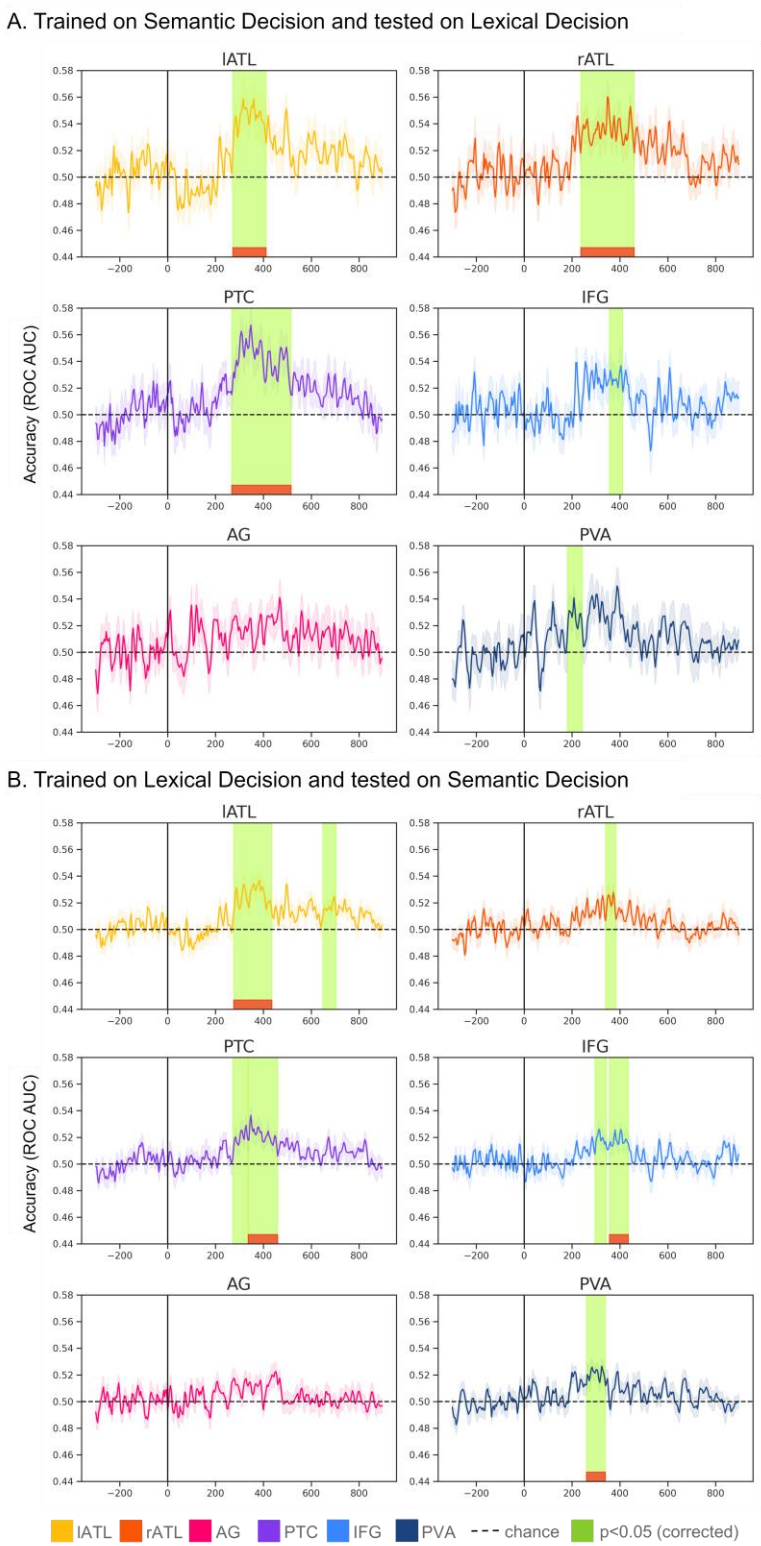


Figure 5. Cross-task semantic category decodability. For all plots, the shadowed area represents the standard error of the mean (across participants). Time points highlighted in green are times when cluster-based permutation correction test revealed a significant cluster (α=0.05). In red, are clusters that are still significant after Bonferroni correction (across ROIs). ***A.*** *Individual ROIs cross-decoding when the model was trained on SD trials and tested on LD trials* ***B.*** *Individual ROIs cross-decoding when the model was trained on LD trials and tested on SD trials.*

As a final step, we wanted to determine whether the decoding performance was driven by any particular semantic word categories, so we examined the confusion matrices of each significant temporal cluster. To obtain a confusion matrix for each cluster, we obtained the model's predicted class for each time point, separately for each participant. We then fitted a (normalised) confusion matrix for each point and considered the average values across time for each cluster. Figure 6 shows the confusion matrices for all the temporal clusters that were significantly above chance across participants in the semantic category classification, for each task/ROI separately. The confusion matrices indicate that abstract words are the most differentiable (i.e., relatively higher accuracy) from the other classes, which were all concrete.

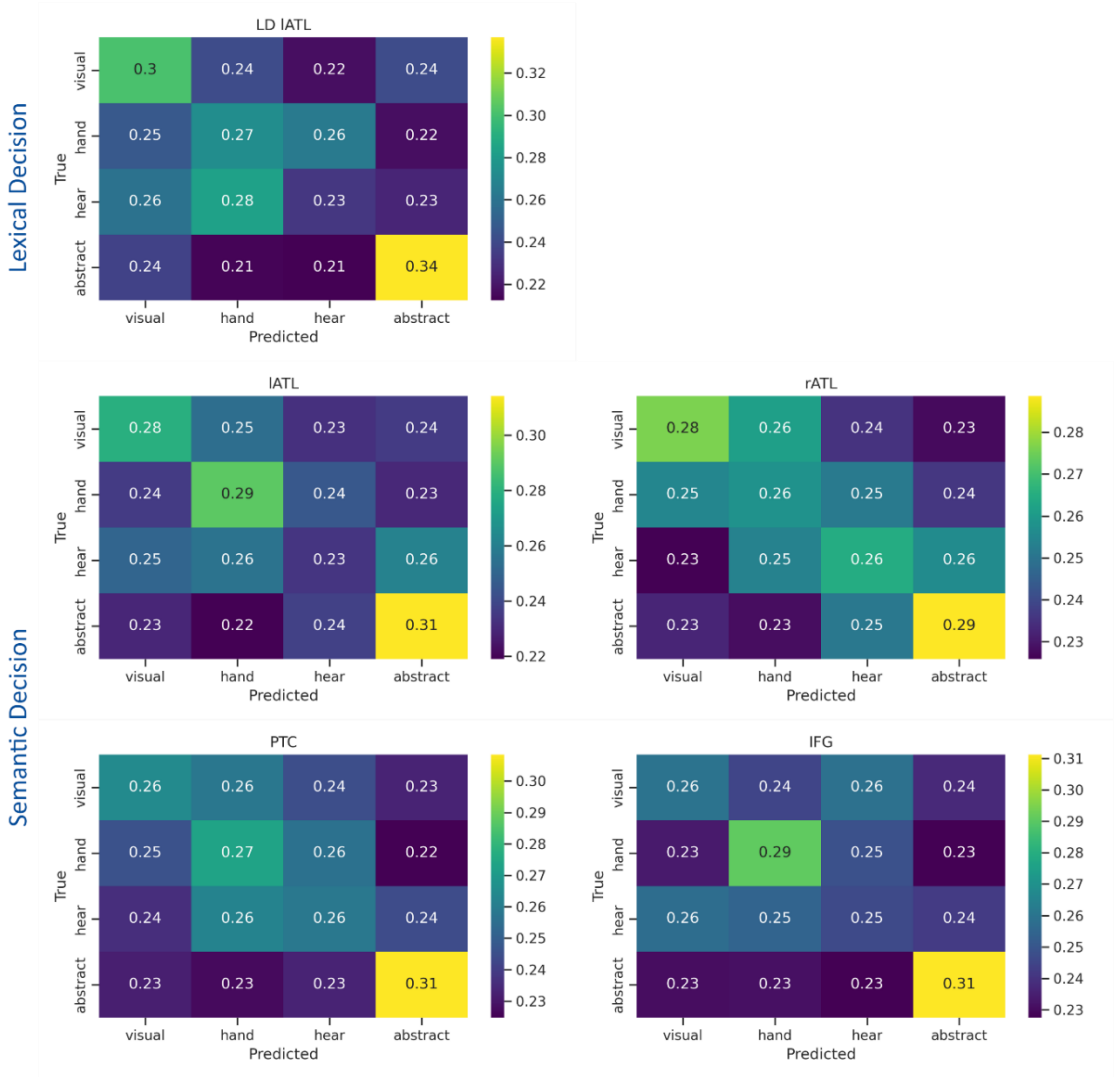


Figure 6. Confusion Matrices for the decoding of semantic word categories. Each cell represents the temporally-averaged probability of a predicted word category, given the true word category, separately for each of the significant temporal clusters found in the semantic category classification. Smaller values indicate lower probabilities. In all cells, abstract words were the most accurately predicted category, compared to the other (concrete) words.

Combined ROIs Patterns

Similarly to the individual ROI case, accuracy was reliably above chance only in the SD but not LD task. In SD, one cluster extended between 328-412ms and a second cluster approximately between 524-568ms. However, also in this case, statistical analysis revealed no difference between SD and LD models. Decoding accuracy for individual ROIs as well as for combined ROIs was quite low (below 0.55) compared to our results in Figure 3. As a result, the activation patterns for the Combined-ROI analysis appeared to be less informative in this case. For both LD and SD these patterns showed an almost linear, drift-like increase over time following the baseline period, with a small peak at about 100ms for PVA. This confirms the assertion of Haufe et al. (2014) that classifier weights and activation patterns are less interpretable when decoding accuracies are low.

Discussion

Our study had two main goals: First, we wanted to characterise the spatiotemporal dynamics within the semantic brain network using multivariate information decoding; and second, we compared two different approaches to source space decoding, based on individual ROI and combined ROI data.

Logistic regression models performed very well at classifying task and showed high accuracy in decoding LD vs SD trials in all ROIs from before 100ms. This is consistent with previous studies that reported effects of task on visual word processing from early latencies (around 100ms, Chen et al., 2015) and the evoked and functional connectivity results obtained from the same data (Rahimi et al., 2022). However, these effects between LD and SD may not be surprising, since these tasks differ in several ways, such as the semantic depth of processing, task difficulty and attentional demands (as reflected in response times), and stimuli were presented in different blocks. More interestingly, the classifier could distinguish with high accuracy between different SD blocks, which reflected more subtle task demands as they only differed with respect to the semantic target category for the response. In all task classifications, all ROIs showed an early increase in decoding accuracy from about 50 ms which was followed by a plateau, and a slow decline towards the end of the epoch. This plateau indicates that for a sustained period, the whole semantic network carries information about the task that is being performed.

Also, we were able to decode single-word semantic category across the whole semantic network (left and right Anterior Temporal Lobe, left Posterior Temporal Cortex, and left Inferior Frontal Gyrus) in the semantic decision task, and in the left Anterior Temporal Lobe in the lexical decision task. This indicates that semantic information is represented across all

the semantic network, at least when performing a semantically engaging task. However, statistical tests revealed no significant differences between the two tasks in the direct comparison in any ROI. Thus, word category information may be present across the semantic network in both tasks, but only reach significance in the SD task. This was confirmed by our additional cross-task decoding analysis: when training the classifier on semantic decision and testing on lexical decision trials, and vice versa, we could successfully decode stimulus features to a similar degree as in individual tasks. Thus, while a previous study revealed significant differences between tasks in evoked activity and functional connectivity (Rahimi et al., 2022), our current decoding results show that when it comes to semantic feature representation there is a higher degree of similarity across tasks. Altogether, this supports the view that early visual word recognition processes are flexible, i.e. they are modulated by task demands, but semantic information is processed even if not explicitly required by the task (Chen et al., 2015; Evans et al., 2012).

Only Angular Gyrus and Primary Visual Cortex did not produce above-chance decoding accuracy for semantic word category, neither in LD nor in SD blocks. This confirms previous results that early brain activity in the temporal lobes, but not in angular gyrus, reflects semantic word features (Farahibozorg et al., 2022). It also supports the view that AG does not support semantic representation per se, but is involved in more spatiotemporally extended semantic processing (Humphreys et al., 2021). The absence of PVA effects in the category decoding analysis indicates that perceptual properties of the stimuli were not influencing the classification performance, as our stimuli were well matched for several relevant psycholinguistic variables. Thus, while for pictorial stimuli semantic category is often confounded by visual stimulus features (e.g., natural concepts being more “curvy” and artefacts being more “edgy”), our word recognition results are most likely due to semantic stimulus features.

Finally, we tested whether and how different semantic features affected the decoding performance in different regions. We addressed this question by computing confusion matrices for the decoding of different word categories against each other. This revealed that concrete word categories were more likely to be confused with each other (mostly indistinguishable) and that abstract words were driving the decoding performance across all regions that showed above-chance performance (i.e., all the semantic network in the SD task). This confirms that information about a word’s concreteness is represented in the activation patterns within the semantic brain network from early latencies, as demonstrated in previous studies at a univariate level (e.g., Farahibozorg et al., 2022). Also, this indicates that multivariate brain patterns reflect semantic similarities (i.e., more similar concepts have similar representations) and parallel fMRI results (Kriegeskorte et al., 2008).

A methodological objective of our study was to compare two approaches to source-space decoding. We looked at both task classification and stimulus classification using decoding per ROI and decoding across ROI. While we found reliable task decoding using both approaches, for the semantic category decoding only the individual-ROI approach yielded reliable results. This suggests that when the effect of interest is distributed across brain regions, then the across-ROI approach is more informative, but when the effect of interest is distributed more sparsely over space and time, i.e. if it is present only at certain latencies or regions, then the per-ROI approach is more sensitive.

Decoding performance was high both for general (LD-SD) and more subtle (SD-SD) tasks, across all our ROIs. Although in both cases accuracy varied across ROIs (e.g., relatively higher in PVA and AG compared to semantic ROIs in the LD-SD task), it is not straightforward how to interpret this as evidence of different degrees of information contained in each ROI, as different algorithms and parameters will likely influence the accuracy score. For example, this difference could just reflect the “visibility” of a region by the EEG/MEG sensors. Hence, the individual-ROI approach was not very informative concerning the time course of the effects (contrary to the semantic category decoding), as all ROIs showed a similar timecourse (rapid increase in performance, followed by a plateau, and slow decline). However, more information was available when observing the activation patterns of the model that combined across-ROIs information. For example, we observed that while the relative contribution of visual areas starts early, semantic regions take longer to reach their peak, consistent with univariate results of a posterior-to-anterior sweep of information which is observed in evoked responses (Marinkovic et al., 2003; Rahimi et al., 2022). Interestingly, we found no obvious difference between ROIs relevant for the LD-SD and SD-SD classifications.

The only region that showed significant decoding accuracy across all our analyses was the left ATL, with the caveat that the differences in stimulus decoding accuracy did not differ statistically between the lexical and semantic decision task. Nevertheless, we found involvement of those regions that have been put forward as the core semantic network (Lambon Ralph et al., 2017). More specifically, it has been suggested that IFG exerts semantic control function via PTC onto ATL (Jackson et al., 2021). Our results are consistent with this framework. Information regarding the semantic properties of a word seems to be spread across the core semantic network, but probably not outside (at least not in PVA and AG). This semantic information is at least partially stable across tasks. Future studies should investigate the information flow based on multivariate pattern as well as connectivity analyses in more detail. The amount of multivariate methods available for evaluating semantic computations from neuroimaging data is rapidly increasing (see Frisby et al., 2023 for a review). Furthermore, methods that characterise brain connectivity based on

multidimensional relationships, or pattern-to-pattern transformations, have recently become available (Basti et al., 2020; Rahimi, Jackson, Farahibozorg, et al., 2023; Rahimi, Jackson, & Hauk, 2023).

In conclusion, our results demonstrate that EEG/MEG source space activity contains rich information about stimulus and task features in written word recognition, which will be essential to unravel the complex dynamics in the semantic brain network.

Acknowledgements

This work was supported by intramural funding from the Medical Research Council UK (MC_UU_00005/18), and by Cambridge ESRC DTP and Cambridge European & Newnham College scholarships awarded to F.M. For the purpose of open access, the author has applied a CC BY public copy-right license to any Author Accepted Manuscript version arising from this submission. We thank Rezvan Farahibozorg for data collection and information on the dataset. We would like to thank also Setareh Rahimi, Pranay Yadav, and Kevin Campion for helpful discussions and support during the preparation of this manuscript.

References

Basti, A., Nili, H., Hauk, O., Marzetti, L., & Henson, R. N. (2020). Multi-dimensional connectivity: A conceptual and mathematical review. *NeuroImage*, *221*, 117179. https://doi.org/10.1016/j.neuroimage.2020.117179

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic Bulletin & Review*, *23*(4), 1096–1108. https://doi.org/10.3758/s13423-015-0909-1

Chan, A. M., Baker, J. M., Eskandar, E., Schomer, D., Ulbert, I., Marinkovic, K., Cash, S. S., & Halgren, E. (2011). First-Pass Selectivity for Semantic Categories in Human Anteroventral Temporal Lobe. *Journal of Neuroscience*, *31*(49), 18119–18129. https://doi.org/10.1523/JNEUROSCI.3122-11.2011

Chen, Y., Davis, M. H., Pulvermüller, F., & Hauk, O. (2015). Early Visual Word Processing Is Flexible: Evidence from Spatiotemporal Brain Dynamics. *Journal of Cognitive Neuroscience*, *27*(9), 1738–1751. https://doi.org/10.1162/jocn_a_00815

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Dhond, R. P., Witzel, T., Dale, A. M., & Halgren, E. (2007). Spatiotemporal cortical dynamics underlying abstract and concrete word reading. *Human Brain Mapping*, *28*(4), 355–362. https://doi.org/10.1002/hbm.20282

Evans, G. A. L., Lambon Ralph, M. A., & Woollams, A. M. (2012). What's in a word? A parametric study of semantic influences on visual word recognition. *Psychonomic Bulletin & Review*, *19*(2), 325–331. https://doi.org/10.3758/s13423-011-0213-7

Farahibozorg, S.-R. (2018). *Uncovering Dynamic Semantic Networks in the Brain Using Novel Approaches for EEG/MEG Connectome Reconstruction*. https://www.repository.cam.ac.uk/handle/1810/278024

Farahibozorg, S.-R., Henson, R. N., Woollams, A. M., & Hauk, O. (2022). Distinct Roles for the Anterior Temporal Lobe and Angular Gyrus in the Spatiotemporal Cortical

Semantic Network. *Cerebral Cortex*, bhab501. https://doi.org/10.1093/cercor/bhab501

Frisby, S. L., Halai, A. D., Cox, C. R., Lambon Ralph, M. A., & Rogers, T. T. (2023). Decoding semantic representations in mind and brain. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2022.12.006

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. https://doi.org/10.1038/nature18933

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. S. (2013). MEG and EEG Data Analysis with MNE-Python. *Frontiers in Neuroscience*, *7*(267), 1–13. https://doi.org/10.3389/fnins.2013.00267

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex | Science. *Science*, *293*. https://www.science.org/doi/full/10.1126/science.1063736

Humphreys, G. F., Lambon Ralph, M. A., & Simons, J. S. (2021). A Unifying Account of Angular Gyrus Contributions to Episodic and Semantic Cognition. *Trends in Neurosciences*, *44*(6), 452–463. https://doi.org/10.1016/j.tins.2021.01.006

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Isik, L., Meyers, E. M., Lelbo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of Neurophysiology*, *111*, 91–102. https://doi.org/10.1152/jn.00394.2013

Jackson, R. L. (2021). The neural correlates of semantic control revisited. *NeuroImage*, *224*, 117444. https://doi.org/10.1016/j.neuroimage.2020.117444

Jackson, R. L., Rogers, T. T., & Lambon Ralph, M. A. (2021). Reverse-engineering the cortical architecture for controlled semantic cognition. *Nature Human Behaviour*, *5*(6), 774–786. https://doi.org/10.1038/s41562-020-01034-z

Jefferies, E. (2013). The neural basis of semantic cognition: Converging evidence from neuropsychology, neuroimaging and TMS. *Cortex*, *49*(3), 611–625. https://doi.org/10.1016/j.cortex.2012.10.008

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, *116*(43), 21854–21863. https://doi.org/10.1073/pnas.1905544116

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal Dynamics of Modality-Specific and Supramodal Word Processing. *Neuron*, *38*(3), 487–497. https://doi.org/10.1016/S0896-6273(03)00197-1

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Molins, A., Stufflebeam, S. M., Brown, E. N., & Hämäläinen, M. S. (2008). Quantification of the benefit from integrating MEG and EEG data in minimum ℓ2-norm estimation. *NeuroImage*, *42*(3), 1069–1077. https://doi.org/10.1016/j.neuroimage.2008.05.064

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987. https://doi.org/10.1038/nrn2277

Rahimi, S., Farahibozorg, S.-R., Jackson, R., & Hauk, O. (2022). Task modulation of spatiotemporal dynamics in semantic brain networks: An EEG/MEG study. *NeuroImage*, *246*, 118768. https://doi.org/10.1016/j.neuroimage.2021.118768

Rahimi, S., Jackson, R., Farahibozorg, S.-R., & Hauk, O. (2023). Time-Lagged Multidimensional Pattern Connectivity (TL-MDPC): An EEG/MEG pattern transformation based functional connectivity metric. *NeuroImage*, *270*, 119958. https://doi.org/10.1016/j.neuroimage.2023.119958

Rahimi, S., Jackson, R., & Hauk, O. (2023). *Identifying nonlinear Functional Connectivity with EEG/MEG using Nonlinear Time-Lagged Multidimensional Pattern Connectivity (nTL-MDPC)* [Preprint]. Neuroscience. https://doi.org/10.1101/2023.01.19.524690

Rogers, T. T., Cox, C. R., Lu, Q., Shimotake, A., Kikuchi, T., Kunieda, T., Miyamoto, S., Takahashi, R., Ikeda, A., Matsumoto, R., & Lambon Ralph, M. A. (2021). Evidence for a deep, distributed and dynamic code for animacy in human ventral anterior temporal cortex. *ELife*, *10*, e66276. https://doi.org/10.7554/eLife.66276