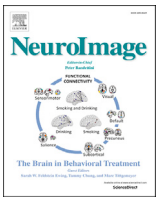
NeuroImage 219 (2020) 116913

Contents lists available at ScienceDirect

NeuroImage

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



Spatiotemporal properties of the neural representation of conceptual content for words and pictures – an MEG study

Giuliano Giari [a], Elisa Leonardelli [a], Yuan Tao [b], Mayara Machado [c], Scott L. Fairhall [a,*]

[a] Center for Mind/Brain Science, University of Trento, Trento, Italy
[b] Department of Cognitive Science, Johns Hopkins University, Baltimore, United States
[c] Max Planck Institute for Human Development, Berlin, Germany

A B S T R A C T

The entwined nature of perceptual and conceptual processes renders an understanding of the interplay between perceptual recognition and conceptual access a continuing challenge. Here, to disentangle perceptual and conceptual processing in the brain, we combine magnetoencephalography (MEG), picture and word presentation and representational similarity analysis (RSA). We replicate previous findings of early and robust sensitivity to semantic distances between objects presented as pictures and show earlier (~105 msec), but not later, representations can be accounted for by contemporary computer models of visual similarity (AlexNet). Conceptual content for word stimuli is reliably present in two temporal clusters, the first ranging from 230 to 335 msec, the second from 360 to 585 msec. The time-course of picture induced semantic content and the spatial location of conceptual representation were highly convergent, and the spatial distribution of both differed from that of words. While this may reflect differences in picture and word induced conceptual access, this underscores potential confounds in visual perceptual and conceptual processing. On the other hand, using the stringent criterion that neural and conceptual spaces must align, the robust representation of semantic content by 230–240 msec for visually unconfounded word stimuli significantly advances estimates of the timeline of semantic access and its orthographic and lexical precursors.

1. Introduction

Successful interaction with the environment requires both the perception of objects and access to their meaning. At the neural level, the entwined nature of these two processes renders an understanding of the interplay between perceptual recognition and conceptual access a continuing challenge. The neural substrates of visual object perception are known to be highly sensitive to different semantic classes or 'object-categories' (e.g. faces, Kanwisher et al., 1997; places, Epstein and Kanwisher, 1998; body parts, Downing et al., 2001). However, visual features strongly covary with object category and the selective activation of visual cortex may reflect variations in visual rather than semantic processing. Model-driven analytical techniques such as representational similarity analysis (RSA; Kriegeskorte et al., 2008), which allow the assessment of concordance between neural and cognitive representational spaces, might have potential to disentangle the specific contribution of each process. In visual perception, functional magnetic resonance imaging (fMRI) indexed neural representational distances between object categories conforms to the subjective similarity ratings in object-selective ventral visual cortex and to Hmax (Serre et al., 2007) computer model of vision in V1 (Connolly et al., 2012). RSA has further shown that MEG representational space of visually presented object evident at 90 msec conform to fMRI representational spaces in V1, while MEG representational spaces at 130 msec conform to fMRI representational spaces in ventral visual cortex (Cichy et al., 2014).

However in both these cases, representation in ventral visual cortex may also be attributable to visual features (Kaiser et al., 2016). Further attempts have been made to resolve this ambiguity using competing models to disentangle visual from semantic features. Clarke et al. (2015) demonstrated that semantic-feature models explained variance beyond that accounted for by Hmax visual models from 110 msec post stimulus onset (see also Bankson et al., 2018). However, also in this context, dissociating perceptual from conceptual contributions to content sensitivity is an ill posed problem, contingent on the capacity of visual models to capture all perceptual contributions. As conceptual representations can be accessed not only through pictures but also words, word offer a perceptually unconfounded method to cue conceptual representations. A growing number of fMRI studies that have employed word, or cross modal (word/picture) stimuli have observed sensitivity to semantic content in a much-reduced subset of brain regions, largely distinct from those showing sensitivity to object category when viewing pictures (Bruffaerts et al., 2013; Fairhall and Caramazza, 2013; Liuzzi et al., 2017; Simanova et al., 2014). These studies provide insight into the locus of conceptual representation in the brain but cannot provide insight into the temporal properties of access.

* Corresponding author.
E-mail address: scott.fairhall@unitn.it (S.L. Fairhall).

https://doi.org/10.1016/j.neuroimage.2020.116913
Received 3 September 2019; Received in revised form 26 April 2020; Accepted 1 May 2020
Available online 7 May 2020

The goal of the present study is to identify the differential time course of conceptual and perceptual processing. We used magnetoencephalography with RSA to assess the relationship between conceptual representational spaces and neural representational spaces during picture presentation. Then, similar to previous research, we controlled for visual similarity using a contemporary neural network models of vision, AlexNet (Krizhevsky et al., 2012). Finally, and critically, we examine the relationship between conceptual similarity and neural similarity spaces elicited by words, where perceptual features are dissociated from meaning, providing an unclouded view of conceptual access as well as a baseline for the extent to which current computer models of vision can fully account for human perceptual processes.

2. Materials & methods

2.1. Participants

Twenty-five right-handed native Italian speakers underwent an MEG recording session (14 male; mean age 25.1 years, SD = 5.36). All participants had normal or corrected-to- normal vision and reported no history of neurological disorders. Prior to recording they gave written informed consent to participate in the experiment. All procedures were approved by the ethical committee of the University of Trento.

2.2. Stimuli

Stimuli consisted of written Italian words and corresponding pictures of exemplars from five semantic categories: fruits, tools, clothes, mammals and birds. Each category was composed of 32 exemplars for a total of 160 exemplars. Word length did not differ between categories ($F(4,155) = 1.83$, $p = .126$) and was uncorrelated with the conceptual model ($r = -0.004$). Likewise, word frequency (Lyding, 2014) did not differ across categories ($F(4,155) = 1.02$, $p = .40$) or correlate with the conceptual model ($r = 0.03$). In picture presentation trials, pictures were randomly selected from 5 different images of that object.

2.3. Procedure

The experiment was composed of a total of seven runs, the first two using word stimuli while the other five pictures. A greater number of picture runs were presented as the availability of differing exemplars was anticipated to engage participants more than the presentation of the necessarily identical stimulus in the word runs. While this different number of runs for words and pictures may affect variability across participants, it will not influence the magnitude of the similarity effect. Two subjects completed only three runs of pictures. Each run consisted in the presentation of all the 160 exemplars, divided in 20 blocks of 8 stimuli from one category. Blocks were preceded by a word cue indicating the upcoming category (750 msec) followed by a fixation cross (750 msec). Block order was pseudo randomised within runs. Trials consisted of a stimulus (word or picture) presented for 400 msec followed by a fixation cross 2100 ms. On each trial, participants rated the typicality of the exemplar as a member of the category on a scale from 1 (very typical) to 4 (not typical) via a bi-manual VPixx button box (VPixx technologies, Canada). Having blocks of stimuli from the same category was important for subjects to perform the behavioral typicality task. Presentation was controlled using Psychtoolbox (Brainard, 1997). Each image was back-projected with a VPixx PROPixx projector at the centre of a translucent screen placed 120 cm from the eyes of the participants. Refresh rate of the screen was 120 frames per second. Timing was verified with a photodiode.

2.4. MEG data acquisition

Continuous MEG data were recorded at the Center for Mind and Brain Sciences of the University of Trento using Elekta Neuromag 306 MEG system (Elekta, Helsinki, Finland), composed of 102 magnetometers and 204 planar gradiometers, placed in a magnetically shielded room (AK3B, Vakuumschmelze, Hanau, Germany). Prior to the experiment, participants' head shape was recorded using a 3D digitizer (Fastrak Polhemus, Inc., Colchester, VA, USA), including the fiducial points (nasion, left and right periaricular sites) and the position of five coils (one on the left and right mastoid, three on the forehead). Before each run head position within the MEG helmet was recorded by inducing a non-invasive current though the five coils. Data were collected at 1000 Hz and hardware filters were adjusted to bandpass the MEG signal in the frequency range of 0.01–330 Hz.

2.5. MEG data preprocessing

Raw MEG data were visually inspected to identify bad channels, according to their overall noise level and signal jumps. An average of 7 channels per run were discarded. After this procedure 4 subjects were discarded due to the high number of bad channels (>12). For each subject a reference run was calculated as the one that minimizes the distance in head position. This was then used to realign all other runs and thus have a uniquely defined head position. Realignment and bad channel interpolation were conducted through the temporal signal space separation method (TSSS; Taulu and Simola, 2006) as implemented in the MaxFilter software, thus reducing dimensionality of the data by retaining only components that are not correlated to noise. Additional preprocessing was then conducted using the Fieldtrip package (Oostenveld et al., 2011). Maxfiltered data were low-pass filtered at 80 Hz and high-pass filtered at 0.8 Hz to remove slow drifts in the signal. Although recent evidence shows that high pass filtering has moderate effects on the temporal profile of classifiers performance (van Driel et al., 2019), no such effect has been reported in the context of RSA analyses. In order to understand its impact, we examined the time course of RSA results on a single random subject with and without filtering. The observed results show a diminished effect as the filter cut-off increases, thus the current filter settings should not lead to any false positives. A notch filter was applied at 50 Hz to remove line noise. Photodiode was used to correct for timings recorded by the MEG system. These corrected timings were used to segment epochs of 1 s (from −0.2 to 0.8 s), resulting in 160 trials per run. Data were then down-sampled to 200 Hz to increase signal-to-noise ratio in the multivariate analysis and to reduce computational time (Grootswagers et al., 2017). Runs were concatenated and visually inspected. Using the function ft_rejectvisual with the summary method, single-trial variance over time was computed and the mean over channels was plotted to identify trials that exceeded the general trend of the individual subject. An average of 19 trials per subjects were discarded. Each trial was baseline corrected, subtracting the mean activity in the baseline window (from -.2 to 0 s) from the subsequent time points. To take into account the different measurement units of magnetometer (T) and gradiometers (T/m) sensors, the signal recorded by magnetometers was multiplied by a factor of 17 corresponding to their distance in mm from the gradiometers, so to have them on the same scale and have a comparable influence on correlation value calculations.

2.6. Representational similarity analysis

In the present study we used Representational Similarity Analysis (RSA; Kriegeskorte, 2008) to compare the representational structure between the MEG data and conceptual dissimilarity ratings (separately for pictures and words). We then used a computational model of the visual system to assess the influence of visual features on picture-based representational spaces. The goal of RSA is to compare the representational vector spaces associated with the different experimental conditions to obtain a measure of their (dis)similarity. This results in a representational dissimilarity matrix (RDM) containing in each cell the pairwise distance value between different conditions. The main advantage of RSA is the possibility to compare RDMs obtained from different

kinds of data. These in fact reflect only the representational structure, not preserving the specific geometry of the features used in its construction. Neural RDMs can then be compared to models to test the hypothesis that brain activity is shaped accordingly to the model.

2.6.1. Neural RDMs

Preprocessed data were mean-centred by removing the mean across-trial pattern (Diedrichsen and Kriegeskorte, 2017). At each time point, we correlated the MEG activity between trial pairs, separately for the picture and word modality. This results in a distance value (1- Pearson correlation) that indicates the dissimilarity between trial pairs according to brain activity. By repeating this procedure for each trial pair we constructed RDMs that account for the whole representational space. Individual trials were used as input to the RDM calculation. To calculate the time-point by time-point RDMs, the vector for the 306 sensors was concatenated with those of the two preceding and the two succeeding time points, as implemented in CoSMoMVPA. This resulted in a vector length of 1530 features reflecting brain activity spanning 20 ms. These were then correlated to conceptual similarity judgements (pictures and words) and to conceptual similarity judgements while removing visual similarity (pictures only).

2.6.2. Conceptual RDMs

An independent sample of 10 subjects took part in a separate behavioural experiment adapted from the Multi-Arrangement Method (Kriegeskorte and Mur, 2012). Briefly, subjects positioned pictures of items in a 2-D space, with the mouse drag-and-drop, according to overall conceptual similarity. The pairwise distance between items was then calculated as to obtain a conceptual RDM. This conceptual model was restructured for each participant to match the actual trials' label and correlated with the neural RDMs produced by pictures and words at each time point.

To characterize the information represented in our conceptual model we created a categorical model based on 1s and 0s (1-correlation) for different and similar categories, respectively. The estimated correlation with the conceptual model was r = 0.79, indicating that the categorical structure of individual concepts is represented in the model, but more fine-grained differences persist.

2.6.3. Visual RDMs

Convolutional neural networks (CNNs) are considered a reliable model of the human isual system with their features reflecting its gradient in complexity (Cichy et al., 2016; Güçlü and van Gerven, 2015). In the present study we used a pretrained AlexNet architecture (Krizhevsky et al., 2012) as implemented in Matlab (Fig. 1B. NN architecture was adapted from "Typical NN architecture" by Aphex34 used under BY-SA 4.0). We gave as input to the neural network all the images used for this study and extracted the activations at four specific layers (conv1, conv3, conv5, fc8). Convolutional layers (conv) activation is a feature map representing progressively bigger local image features through arbitrary units. The final fully-connected (fc8) activation is instead a vector with 1000 entries, each one representing the probability of the input image to be classified as one of 1000 classes. For each subject then, we matched the activations of each layer to the actually presented images and calculated the pairwise dissimilarity between feature vectors as 1-Spearman correlation. We used these RDMs as regressors in a partial correlation approach, thus by correlating the pictures MEG RDM to the conceptual similarity RDM while removing the influence of visual similarity modelled through these RDMs. Visual RDMs have a mean (between-subject) correlation with the conceptual RDM of r = 0.19, 0.24,

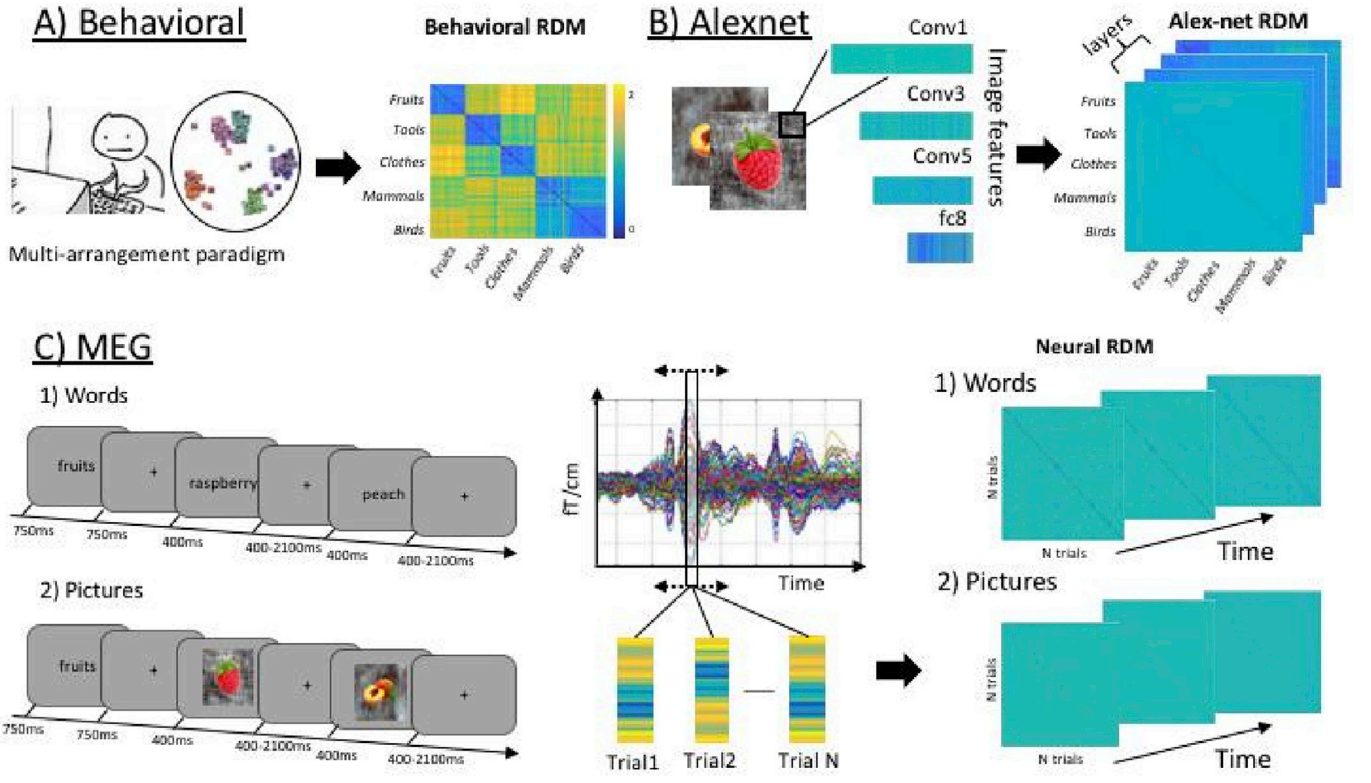


Fig. 1. A) Behavioural experiment with the multi-arrangement paradigm was conducted to obtain a measure of conceptual dissimilarity for 160 exemplars from 5 semantic categories. These were averaged across participants (N = 10) to create a conceptual similarity RDM. B) For each image, we extracted activations at four selected layers of a pretrained AlexNet architecture and constructed visual similarity RDMs. C) In an MEG-experiment (N = 21) the same exemplars were presented first as words (2 runs) and then as pictures (5 runs). Each run was composed of 20 category-specific blocks in which subjects had to rate typicality of 8 exemplars as a member of the instructed category in a scale 1–4. Preprocessed MEG trials were vectorized to calculate pairwise distance resulting in a neural RDM at each time point, separately for pictures and words. Three separate RSAs were performed: one for pictures, one for pictures controlling for visual similarity and one for words.

0.25 and 0.47 for conv1, conv3, conv5 and fc8. respectively.

2.6.4. Searchlight analysis

To identify the spatially localised sources of information content, we ran a searchlight analysis over all channels in the four time windows showing maximal information content in the picture condition. Considering only planar gradiometers, we averaged data ±10 msec around 180, 280, 365 and 540 msec to perform the same RSA analysis described in previous sections on pictures, pictures controlled for visual similarity and words. For each sensor we selected the closest 24 planar channels (i.e. 12 gradiometer pairs) to be included as neighbours. The result of these analysis is a topography of correlation values that indicates the extent to which each sensor (including its neighbours) is correlated to our semantic model.

2.7. Multivariate pattern analysis

Multivariate pattern analysis (MVPA, Haxby et al., 2014) exploits machine learning algorithms to solve a supervised classification problem. A subset of the data is used to find a linear boundary between the multivariate neural activation patterns that best distinguishes two conditions of interest (training), whose reliability is tested on never before seen data. An accuracy value is returned indicating how well learned features are able to predict the class of unlabelled data. An above chance accuracy then indicates that some features are shared between data partitions and thus informative for the class distinction.

We used MVPA to investigate the extent to which the informational content represented in the MEG sensor array is shared across picture and word modalities. Specifically, we trained a Linear Discriminant Analysis (LDA) classifier to distinguish between object-categories. This procedure was reduced to a two-class classification problem by repeating the analysis for all possible combinations of category pairs. Importantly, the classifier was trained on one modality and tested on the other and vice versa. To avoid overfitting we implemented a leave-one-chunk-out cross validation approach. We partitioned the data in 8 chunks and repeated the analysis until all chunks were used both as training and as testing set.

To take into account possible delays between the processing time course of the two modalities (Leonardelli et al., 2019) we used the temporal generalization approach (King and Dehaene, 2014). The training and testing procedure was thus repeated for all possible time point combinations. The final result is a matrix of decoding accuracies indicating the time points in which the informational content that allows the distinction between object-categories is shared across modalities.

2.8. Statistical analysis

RSA results were Fisher transformed prior to statistical analysis to have normally distributed values. Significance was assessed using one-tailed cluster-based permutation test. Following the approach described by Stelzer et al. (2013), 50 null distributions for each subject were created by randomly shuffling condition labels and repeating the RSA analysis with the resulting model. These random RSA time courses were combined across subjects to create 1000000 null distributions against which the data were statistically compared with a *t*-test at each time point. This approach allowed the construction of a null distribution which accurately reflected temporal correlation in the error term as well as other potential characteristics of the noise distribution. Correction for multiple comparisons was assessed through a cluster-based approach. Consecutive significant ($p < 0.05$) timepoints in the interval from 100 to 600 msec were considered to form clusters, whose probability was given by their position within the null distribution A cluster-based permutation test (Maris and Oostenveld, 2007; 10000 permutations), as implemented in Fieldtrip, was used to identify spatial clusters from the searchlight analysis. A cluster-based permutation approach (10000 permutations) was also used to assess statistical significance of the time generalised, cross-modal MVPA in the time window between 100 and 600 msec.

2.9. Data and code availability statement

Data will be made available on request.

3. Results

3.1. Behavioural

Due to equipment malfunction, behavioural data were not available for four subjects. Mean typicality rating did not differ as a function of modality ($F1,20) = 0.17$, $p = .68$) or category ($F(1.9,37.5) = 2.35$, $p = .113$, Greenhouse-Geisser corrected), although there is a moderate interaction between these two factors ($F(4,80) = 3.49$, $p = .011$). Post-hoc pairwise comparisons between modalities did not reveal the origin of this interaction, even at uncorrected thresholds. Due to equipment asynchrony, absolute RTs were not available. However, relative (mean-centred) were preserved, permitting full statistical analysis. There was no overall effect of modality ($F(1,20) = 0.09$, $p = .77$) but there was a strong effect of category ($F(4,80) = 5.25$, $p = .001$), and a modality by category interaction ($F(4,80) = 6.43$, $p < .0001$). Critically, RT differences did not systematically correlate across participants with the conceptual RDM across subjects for either words ($t = 0.91$) or pictures ($t = 0.36$) and will not influence RSA analyses.

3.2. Neuroconceptual representational similarity analysis

We calculated dissimilarity matrices at each time point of MEG recordings in which subjects were presented with either pictures or words depicting objects from five different semantic categories (fruits, tools, clothes, mammals, birds). We correlated these RDMs with a semantic model obtained from behavioural ratings collected with the multi-arrangement method in a separate experiment. Additionally, for the picture condition we repeated the same analysis controlling for the visual features of each stimuli by regressing out the visual dissimilarity matrices calculated using the activation of four layers of the AlexNet CNN. Informational content was pronounced for picture stimuli, both before and after controlling for visual similarity via the AlexNet model ($p < .00001$, corrected) and the time course of information content was highly similar (Fig. 2). Picture induced information content was characterised by four peaks with their centres at 180, 280, 365 and 540 msec. The sole exception to this pattern of an early peak in visual information at 105 msec which was absent after controlling for visual similarity. For words stimuli, two temporal clusters survived corrections for multiple comparison, one from 230 to 335 msec ($p = .0002$, corrected) and a second from 360 to 585 msec ($p < .00001$, corrected) (see Fig. 3).

3.2.1. Searchlight analysis

To identify spatially localised sources of information content contributing to the timeseries effects, we performed searchlight analyses on local subsets of 24 planar gradiometers within 20 msec time-windows centred peaks observed in the picture condition (see methods). Picture induced information content was detectable at each sensor location. Controlling for visual features did not alter this topographic distribution of information content. Within the first three time-windows, sensors over occipital sites contain the strongest representation of semantic content. In the last time window, information content is spread more evenly across the sensor space, with frontal recording sites contributing comparable information. Formal statistical inference is not reported as circularity between the definition of time-window of interest and the topographic data render this invalid. However, the extent and strength of the effect are such that significance would persist even after correction for the entire timeseries. In the word modality informational content is more precisely localised and survived correction for multiple comparisons across the four time-windows in three of the considered time windows (critical alpha = .0125). The first and second significant cluster of sensors at 280 ($p = .005$) and 365 msec ($p = .006$) are present over occipital and

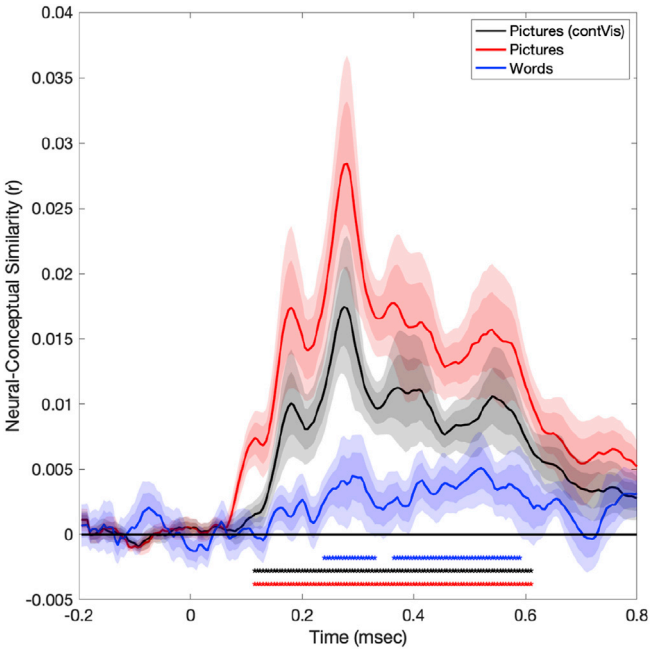


Fig. 2. Time–course of group-level correlations between Conceptual Similarity Ratings and three RDMs derived from MEG-data: pictorial presentation (red), pictorial presentation after partialling out perceptual features through AlexNet (contVis, black) and word presentation (blue). Asterisks indicate clusters that survived multiple comparison correction within the time-window of interest (100 msec–600 msec). All timepoints are significant for the two image-based analyses and two temporal clusters for word stimuli, 230–335 ms ($p = .0002$) and 360 to 585, ($p < .00001$). Darker shaded areas show the standard error of the mean, lighter shaded areas the 95% confidence interval.

temporal sites with the maximal information content present over left occipital temporal sites in the earlier time window and more anterior left temporal sites in the latter window. As was observed for picture stimuli, at 540 msec information content additionally encompasses more frontal recording sites, with semantic information being present broadly across the brain ($p < .001$). A left-lateralised cluster of significant sensors evident in the first time-window for words (lower left topoplot), did not survive correction for multiple comparisons across the four time-windows ($p < .03$).

3.3. Cross-modal MVPA

To detect common conceptual representation produced by word and picture presentation, we performed cross-modal decoding analyses, training an LDA classifier on pictures then testing on words and vice versa (c.f. Fairhall and Caramazza, 2013). As access to conceptual representations may be faster during image presentation than word reading (Leonardelli et al., 2019) we performed time-generalised MVPA (e.g. King and Dehaene, 2014). Specifically, the analysis was aimed at decoding object-category, using a cross-validation strategy where we trained the classifier in one modality and tested in the other. This analysis failed to produce any significant effects ($p = 0.27$).

3.4. Control analyses

To determine the influence of conceptual similarity that extended beyond category, RSA between MEG data and subjective similarity was rerun partialling out the effect of category per se (Figure S1). The results show that neural representations of pictures capture the conceptual relationship between items extending beyond that of object category across all time-points (100–600 msec). On the other hand, word stimuli effects were only observed in two later time windows, from 405 to 460 msec and 485–540 msec. Thus, it is uncertain whether earlier word effects are driven by category or a combination of category and non-categorical semantic similarity.

Given the relatively low-dimensionality underlying our conceptual model compared to richer word corpora, we repeated the same RSA analysis using a semantic model obtained from a word2vec algorithm trained on Italian Wikipedia (Berardi, 2015). Comparable timings of conceptual access are observed (Figure S2).

To evaluate the effectiveness of our visual model, we repeated the searchlight analysis on the first informational content peak (105 msec). While the picture modality already engages the majority of sensors, controlling visual features using AlexNet effectively reduced the

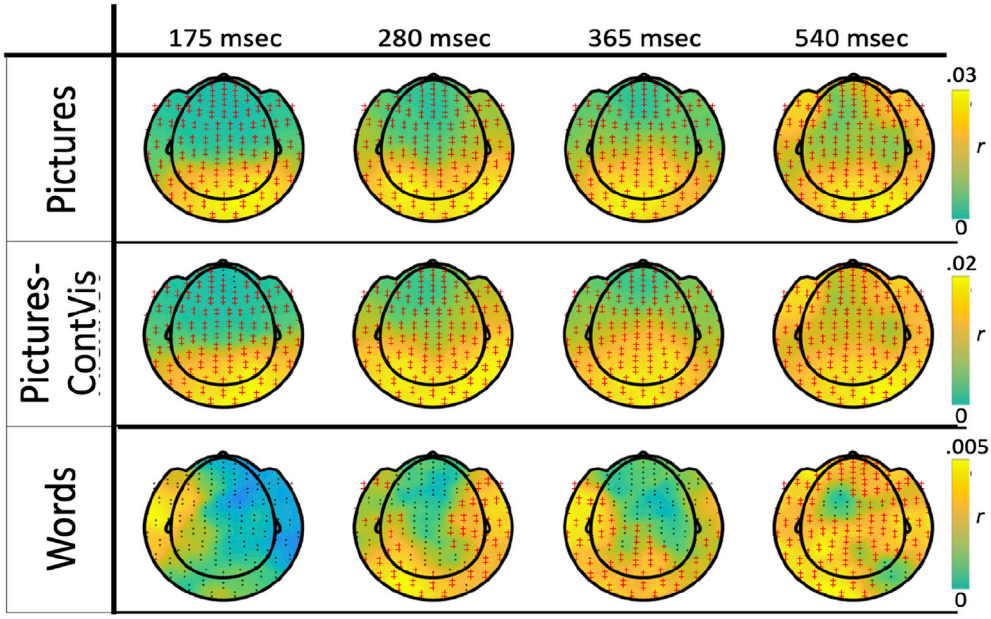


Fig. 3. Topographies of similarities between conceptual and neural similarity spaces. To assess the localisation of information contributing to conceptual content, we performed a searchlight on the gradiometers (radius 12 planar pairs = 24 sensors) at the four peaks of information content identified in the time series analysis for the picture modality. Sensors comprising significant clusters corrected for comparisons across sensors and time-windows are indicated in red.

distribution of localised information content to a topography that is similar to the one observed in the word modality (Figure S3).

To further investigate the contribution of our visual features to the visual neural response, we performed an RSA analysis with separate hierarchical layer as the model RDM, while removing the contribution of the other layers through a partial correlation approach. However, there was a high degree of correlation between RDMs based on the individual layers of Alexnet (r-values: 0.18–0.86) and the contribution of individual layers across timepoints was unclear. Future work on this question should select stimulus sets which have clear and dissociable representational spaces across layers.

Given the particular nature of fc8 activations (class scores) and to safeguard against bias related to the presented object not being one of the fc8 labels, we performed the same partial correlation analysis using the earlier fully connected layer, fc7. While fc8 explained moderately more variance than fc7, the time course of neural conceptual representation was identical when controlling for visual effect via either fc7 or fc8, indicating that the categorical nature of fc8 is not affecting our results.

4. Discussion

To disentangle perceptual and conceptual contributions to object processing, we conducted an MEG study in which stimuli representing exemplars from five semantic categories were presented to participants using two different input modalities, pictures and written words. We used RSA to compare neural representational similarity across time with a model of conceptual similarity constructed using behavioural judgements. Additionally, we used a neural network for visual recognition to control for the influence of perceptual processing in the picture modality. The richness of information content was not uniform across time and followed a similar time course for pictures, and pictures controlled for visual similarity. While an initial peak in information content around 105 msec was present only for pictures when not controlled for visual features, subsequent peaks at 180, 280, 365 and 540 msec were observed both for pictures and pictures when controlled for visual features. Words elicited semantic content across two significant clusters, the first ranging from 235 to 325 msec, the second from 360 to 585 msec.

The first observed peak (105 msec) is present only for picture stimuli. Removing the influence of visual features both via the AlexNet computer model in the picture modality and using words, results in the attenuation of this early peak. Such conceptual-like representations then appear to be attributable to the perception of those visual features that covary with semantic category. As noted previously, representational spaces present in MEG at these early latencies align most closely with fMRI-indexed V1 representations (Cichy et al., 2014), further supporting that cortical responses at these latencies are unrelated to conceptual processing.

In contrast, the second peak in conceptual-neural similarity at 180 msec was not fully explained by the AlexNet model. This result is in line with previous reports (Bankson et al., 2018; Clarke et al., 2015). However, the reliance on pictorial stimuli and the ability of the visual model to capture all features relevant to human perception renders the meaning of this result ambiguous. While there was limited evidence for word-elicited conceptual representation at this time point in both the timeseries and topographic analyses, this effect did not survive correction for multiple comparisons. The timing of this peak coincides with the word selective N170 ERP, a left lateralised component elicited by letter strings more than non-letter stimuli, that is thought to reflect the first stage of orthographic processing (N Bentin et al., 1999; Maurer et al., 2005; Tarkiainen et al., 1999). There is some evidence that more complex processing, orthographic lexical interactions (Coch and Mitra, 2010) or lexicosemantic interactions (Hauk et al., 2006; Kim and Lai, 2012) may occur at this latency. Moreover, there is some evidence that intracranial recording sites may be able to distinguish words denoting animals and man-made objects at similarly early time windows (Chan et al., 2011).

The peak at 280 msec signifies the first robust representation of semantic content for word stimuli and also coincides with the maximal representation of semantic content in the image modality. Word induced evoked potentials within this period show sensitivity to words and pseudowords in comparison to unpronounceable letter strings, are sensitive to word repetition and are thought to reflect further pre-lexical orthographic processing or construction of the word form (Chauncey et al., 2008.; Cohen et al., 2000; Dehaene, 1995; Dufau et al., 2008; Simon et al., 2004). The present finding demonstrates the scope of multivariate pattern analysis in combination with magnetoencephalographic and electrophysiological measures to uncover neural processes normally inaccessible with univariate methodologies. The criteria that neural representational spaces conform to semantic representational spaces provides a strong and valid test that semantic representation underlies this neural response. This result significantly advances the timeline not only for semantic access but for prerequisite orthographic and lexical processing.

Our experimental paradigm might have played a role in the rapidity of this semantic access. Electrophysiological studies of word processing generally employ the random presentation of unrelated words. In the present experiment, stimuli were presented in blocks of related content allowing context to play a role in word processing. Context is an important aspect of skilled reading and can facilitate lexico-semantic processing enabling easier and earlier access (Kutas and Federmeier, 2011). Such contextual facilitations that are common during reading may be lost in paradigms were individual, unrelated words are presented.

A stable significant representation of conceptual content in all the modalities is observed later in the time course with an initial peak in the image modality at 360 msec. This coincides with the N400 potential, a negative deflection culminating at 400 msec in response to semantic violations (Kutas and Federmeier, 2011). The N400 is one of the most well-characterised indices of semantic processing, and has been observed following the presentation of words (Bentin et al., 1985) but also faces (Barrett et al., 1988) and objects (Barrett and Rugg, 1990). It has been demonstrated that its amplitude is greatly influenced by previous processing such as the predictability of a stimulus (Lau et al., 2008) and representations at this time may reflect the integration of semantic information with the working context (Hagoort, 2013). Analysis of the topographic distribution of localised information content shows that, for word stimuli, the relatively constrained occipitotemporal location of information extends to a more distributed representation at the later, 540 msec peak, encompassing occipital, temporal and frontal sites, a pattern also evident for image stimuli. Both this distributed representation of semantic knowledge and the finding that this potential follows from an earlier representation of semantic content further supports a model that the semantic-violation sensitive N400 reflects the integration of incoming semantic information within its working context, rather than indexing the first representation of semantic content.

While there is some consistency in the time course of semantic representation for pictures and words, the topography of information content differed. In the first three time-windows, the searchlight analysis showed that picture representations were detectable at all sensor locations but were strongest over occipital sites. In contrast, word representations were localized in sensors overlying occipitotemporal areas. This indicates that different cortical generators are involved in representing conceptual-like information in word and picture modalities. Only in the latest time-window, 530 msec, did word and picture induced representations share a similar topographic distribution. fMRI studies have demonstrated that large sections of the ventral visual cortex are highly sensitive to object categories and the semantic distance between objects (Connolly et al., 2012; Fairhall and Caramazza, 2013). It is probable that the representations of the presented images in visual cortex dominate the information measured at the sensor level. This may eclipse those more subtle patterns of conceptual representation evident in the word condition underlying neural semantic representation in the 280 and 365 msec time-windows. fMRI studies have implicated the posterior inferior/middle temporal gyrus, precuneus and perirhinal cortex in non-perceptual conceptual representation (Bruffaerts et al., 2013;

Fairhall and Caramazza, 2013; Liuzzi et al., 2017; Simanova et al., 2014). The relationship between these earlier content representations and that occurring at 530 msec to components of the fMRI indexed semantic system remains an important goal for future research.

As in previous studies, we attempted to use visual models to account for the perceptual processing of pictures. The AlexNet model accounted for the image-based response during the earliest time window (~105 msec). Carefully selecting the best model is crucial to fully capture the processing of the human visual system. Although none of them is perfect, AlexNet fully captured early visual responses whereas earlier models such as HMax might not be able to disambiguate them from conceptual-like representations (Clarke et al., 2015). The most parsimonious explanations of the perseveration of representations after this timepoint in the present study is that it arises from a shortfall in the visual model that results in conceptual-like neural representations due to the close relationship between semantics and perceptual features. However, it is also possible that conceptual representations are activated differently when cued by a picture or a word. Viewing a hammer may cue conceptual associations relating to form, affordance or motion, to a greater extent than reading the word. Thus, while we cannot be certain that controlling for visual similarity in picture induced representations only reflect conceptual processing, we cannot be certain that word induced representations reflect all conceptual processing that may occur. In fact, the higher similarity induced by pictorial representations might suggest that some differences persist between modalities even at higher processing stages. In addition, in this study we found no evidence for shared crossmodal representations for word and picture induced conceptual access, in contradistinction to effects seen under comparable conditions in fMRI (Fairhall and Caramazza, 2013). This may be due to the poorer spatial resolution of MEG with strong effects associated with visual perception and localized in visual cortices overwhelming weaker influences of conceptual access at all recorded sensors (unlike fMRI which can spatially segregate 'modal' and 'amodal' cortical sources). This may be further exacerbated by different time-courses in conceptual access, with conceptual access being more rapid during picture presentation which does not necessitate the linguistic processes involved with word reading (Leonardelli et al., 2019). Future MEG studies should take into consideration the spatial overlap and the temporal discrepancy of the effects of interest.

Moreover, the particular influence of a typicality task on the nature of the activated conceptual representation remains uncertain. The task was selected as it encourages access to conceptual representations not constrained to a specific object features (e.g. size, function) in an intuitive and naturalistic manner. Having participants rate the similarity of the specific object to the category in general may have accentuated access to divergent, non-stereotypical features (see also Fairhall, 2020) or, conversely, accentuated those features typical of the category. While it may strongly influence the importance of feature co-occurrence and distinctiveness in the instantiated conceptual representation (Taylor et al., 2011), the specific influence of this on the current task remains uncertain.

In this work, using the stringent criterion that neural and conceptual spaces must align, we observed robust representation of semantic content by 230–240 msec for words. Processing in this time period is usually attributed to orthographic features and the demonstration of the presence of semantic content at this time notably advances estimates of the time course of semantic processing and its orthographic and lexical antecedents. We observe different topographies for information extracted from words or pictures, with information content induced by images being represented most strongly at occipital sensors, both before and after controlling for visual similarity. The spatial location of word and picture induced conceptual representation align only at later time periods (~530 msec) when conceptual information is distributed broadly across recording sites. While words and pictures may cue different types of conceptual access, the differential pattern of information elicited by words and pictures after controlling for visual similarity, argues for caution in the interpretation of the latter.

CRediT authorship contribution statement

Giuliano Giari: Formal analysis, Writing - original draft. **Elisa Leonardelli:** Formal analysis, Writing - review & editing. **Yuan Tao:** Formal analysis. **Mayara Machado:** Investigation. **Scott L. Fairhall:** Conceptualization, Formal analysis, Writing - original draft, Supervision, Funding acquisition.

Acknowledgements

The project was funded by the European Research Council (ERC) grant CRASK - Cortical Representation of Abstract Semantic Knowledge, awarded to Scott Fairhall under the European Union's Horizon 2020 research and innovation program (grant agreement no. 640594).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116913.

References

Bankson, B.B., Hebart, M.N., Groen, I.I.A., Baker, C.I., 2018. The temporal evolution of conceptual object representations revealed through models of behavior, semantics and deep neural networks. Neuroimage 178 (November 2017), 172–182. https://doi.org/10.1016/j.neuroimage.2018.05.037.

Barrett, S.E., Rugg, M.D., 1990. Event-related potentials and the semantic matching of pictures. Brain Cognit. 14 (2), 201–212. https://doi.org/10.1016/0278-2626(90)90029.

Barrett, S.E., Rugg, M.D., Perrett, D.I., 1988. Event-related potentials and the matching of familiar and unfamiliar faces. Neuropsychologia 26 (1), 105–117. https://doi.org/10.1016/0028-3932(88)90034-6.

Bentin, S., McCarthy, G., Wood, C.C., 1985. Event-related potentials, lexical decision and semantic priming. Electroencephalogr. Clin. Neurophysiol. 60 (4), 343–355. https://doi.org/10.1016/0013-4694(85)90008-2.

Berardi, G., Esuli, A., Marcheggiani, D., 2015. Word embeddings go to Italy: a comparison of models and training datasets. In: Italian Information Retrieval Workshop.

Bruffaerts, R., Dupont, P., Peeters, R., Deyne, S. De, Storms, G., Vandenberghe, R., 2013. Similarity of fMRI activity patterns in left perirhinal cortex reflects semantic similarity between words. J. Neurosci. 33 (47), 18597–18607. https://doi.org/10.1523/JNEUROSCI.1548-13.2013.

Chan, A.M., Baker, J.M., Eskandar, E., Schomer, D., Ulbert, I., Marinkovic, K., Halgren, E., 2011. First-pass selectivity for semantic categories in human anteroventral temporal lobe. J. Neurosci. 31 (49), 18119–18129. https://doi.org/10.1523/JNEUROSCI.3122-11.2011.

Chauncey, K., Holcomb, P.J., Grainger, J., Chauncey, K., Holcomb, P.J., Grainger, J., 2008. Language and Cognitive Processes Effects of Stimulus Font and Size on Masked Repetition Priming : an Event- Related Potentials (ERP) Investigation, (June 2013), vols. 37–41. https://doi.org/10.1080/01690960701579839.

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

Cichy, R.M., Khosla, A., Pantazis, D., Torralba, A., Oliva, A., 2016. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Sci. Rep. 6 (June), 1–13. https://doi.org/10.1038/srep27755.

Clarke, A., Devereux, B.J., Randall, B., Tyler, L.K., 2015. Predicting the time course of individual objects with MEG. Cerebr. Cortex 25 (10), 3602–3612. https://doi.org/10.1093/cercor/bhu203.

Coch, D., Mitra, P., 2010. Word and pseudoword superiority effects reflected in the ERP waveform. Brain Res. 1329, 159–174.

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M.A., Michel, F., 2000. The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain 123 (Pt2), 291307.

Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.-C., Haxby, J.V., 2012. The representation of biological classes in the human brain. J. Neurosci. 32 (8), 2608–2618. https://doi.org/10.1523/JNEUROSCI.5547-11.2012.

Dehaene, S., 1995. Evidence for category-specific word processing in the normal human brain. NeuroReport 6 (2), 2153–2157. https://doi.org/10.1097/00001756-199511000-00014.

Diedrichsen, J., Kriegeskorte, N., 2017. Representational models: a common framework for understanding encoding, pattern-component, and representational-similarity analysis. PLoS Comput. Biol. 13 https://doi.org/10.1371/journal.pcbi.1005508.

Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. Science 293 (5539), 2470–2473. https://doi.org/10.1126/science.1063414.

Dufau, S., Grainger, J., Holcomb, P.J., 2008. An ERP investigation of location invariance in masked repetition priming. Cognit. Affect Behav. Neurosci. 8 (2), 222–228. https://doi.org/10.3758/CABN.8.2.222.

Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. Nature 392 (6676), 598–601. https://doi.org/10.1038/33402.

Fairhall, S.L., April 8, 2020. Cross recruitment of domain-selective cortical representations enables flexible semantic knowledge. J. Neurosci. 40 (15) https://doi.org/10.1523/JNEUROSCI.2224-19.2020, 3096–3103.

Fairhall, S.L., Caramazza, A., 2013. Brain regions that represent amodal conceptual knowledge. J. Neurosci. 33 (25), 10552–10558. https://doi.org/10.1523/JNEUROSCI.0051-13.2013.

Grootswagers, T., Wardle, S.G., Carlson, T.A., 2017. Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. J. Cognit. Neurosci. 29 (4), 667–697. https://doi.org/10.1162/jocn.

Güçlü, U., van Gerven, M.A., 2015. Deep neural networks reveal a gradient in the complexity of neural representations across the brain's ventral visual pathway. J. Neurosci. 35 (27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015.

Hagoort, P., 2013. The memory, unification, and control (MUC) model of language. Automaticity. Contr. Lang. Process. 4 (July), 243–270. https://doi.org/10.4324/9780203968512.

Hauk, O., a is, H., ord, Pulvermüller, F., Marslen-Wilson, W.D., 2006. The time course of visual word recognition as revealed by linear regression analysis of ERP data. Neuroimage 30 (4), 1383–1400. https://doi.org/10.1016/j.neuroimage.2005.11.048.

Haxby, James V., Connolly, Andrew C., Swaroop Guntupalli, J., July 8, 2014. Decoding neural representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. 37 (1) https://doi.org/10.1146/annurev-neuro-062012-170325, 435–56.

Kaiser, D., Azzalini, D.C., Peelen, M.V., 2016. Shape-independent object category responses revealed by MEG and fMRI decoding. J. Neurophysiol. https://doi.org/10.1152/jn.01074.2015 jn.01074.2015.

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. : The Official Journal of the Society for Neuroscience 17 (11), 4302–4311. https://doi.org/10.1098/Rstb.2006.1934.

Kim, A., Lai, V., 2012. Rapid Interactions between Lexical Semantic and Word Form Analysis during Word Recognition in Context : Evidence from ERPs, pp. 1104–1112.

King, J.-R., Dehaene, S., April 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cognit. Sci. 18 (4) https://doi.org/10.1016/j.tics.2014.01.002, 203–10.

Kriegeskorte, N., Mur, M., 2012. Inverse MDS: inferring dissimilarity structure from multiple item arrangements. Front. Psychol. 3 (JUL), 1–13. https://doi.org/10.3389/fpsyg.2012.00245.

Kriegeskorte, N., Mur, M., Bandettini, P.A., 2008. Representational similarity analysis – connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2 (November), 1–28. https://doi.org/10.3389/neuro.06.004.2008.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 1–9 https://doi.org/10.1016/j.protcy.2014.09.007.

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

Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (De) constructing the N400. Nat. Rev. Neurosci. 9 (12), 920–933. https://doi.org/10.1038/nrn2532.

Leonardelli, E., Fait, E., Fairhall, S.L., 2019. Temporal dynamics of access to amodal representations of category-level conceptual information. Sci. Rep. 9 (239) https://doi.org/10.1038/s41598-018-37429-2.

Liuzzi, A.G., Bruffaerts, R., Peeters, R., Adamczuk, K., Keuleers, E., De Deyne, S., et al., 2017. Cross-modal representation of spoken and written word meaning in left pars triangularis. Neuroimage 150 (February), 292–307. https://doi.org/10.1016/j.neuroimage.2017.02.032.

Lyding, V, et al., 2014. The paisa'corpus of italian web texts. 9th Web as Corpus Workshop (WaC-9)@ EACL, pp. 36–43.

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

Maurer, U., Brandeis, D., Mccandliss, B.D., 2005. Visual specialization for reading in English revealed by the topography of the N170 ERP response, 12, pp. 1–12. https://doi.org/10.1186/1744-9081-1-13.

N Bentin, S., Mouchetant-Rostaing, Y., Giard, M.H., Echallier, J.F., Pernier, J., 1999. ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. J. Cognit. Neurosci. 11 (3), 235–260. https://doi.org/10.1162/089892999563373.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011. https://doi.org/10.1155/2011/156869.

Serre, T., Oliva, A., Poggio, T., 2007. A Feedforward Architecture Accounts for Rapid Categorization.

Simanova, I., Hagoort, P., Oostenveld, R., Van Gerven, M.A.J., 2014. Modality-independent decoding of semantic information from the human brain. Cerebr. Cortex 24 (2), 426–434. https://doi.org/10.1093/cercor/bhs324.

Simon, G., Bernard, C., Largy, P., Lalonde, R., 2004. Chronometry of visual word recognition during passive and lexical decision tasks: an erp investigation. Int. J. Neurosci. 114, 1401–1432. https://doi.org/10.1080/00207450490476057.

Stelzer, J., Chen, Y., Turner, R., 2013. Statistical inference and multiple testing correction in classification- based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage 65, 69–82. https://doi.org/10.1016/j.neuroimage.2012.09.063. ISSN 1053-8119.

Tarkiainen, A., Helenius, P., Hansen, P.C., Cornelissen, P.L., Salmelin, R., 1999. Dynamics of letter string perception in the human occipitotemporal cortex. Brain 122 (11), 2119–2131. https://doi.org/10.1093/brain/122.11.2119.

Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys. Med. Biol. 51 (7), 1759–1768. https://doi.org/10.1088/0031-9155/51/7/008.

Taylor, K.I., Devereux, B.J., Tyler, L.K., 2011. Conceptual structure: towards an integrated neurocognitive account. Lang. Cognit. Process. 26 (9), 1368–1401. https://doi.org/10.1080/01690965.2011.568227.

van Driel, Joram, Olivers, Christian N.L., Fahrenfort, Johannes J., 2019. High-pass filtering artifacts in ulti ariate lassification of neural time series ata. Biorxi Preprint. Neurosci. (January 26) https://doi.org/10.1101/530220.