bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Context Free and Context-Dependent Conceptual Representation in the Brain

Zhiyao Gao[1], Li Zheng[2], André Gouws[1], Katya Krieger-Redwood[1], Xiuyi Wang[1], Dominika Varga[3], Jonathan Smallwood[4], Elizabeth Jefferies[1]

[1] Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom;

[2] Department of Psychology, University of Arizona, Tucson, AZ 85719, USA

[3] School of Psychology, University of Sussex, Brighton, BN1 9RH, United Kingdom

[4] Department of Psychology, Queen's University, Kingston, ON K7L 3N6, Canada

Abbreviated title:

Context Free and Dependent Conceptual Representation

Acknowledgements

This work was sponsored by the European Research Council (Project ID: 771863 - FLEXSEM Project).

Author contributions

Z.G. and E.J. designed the experiment. K.K.R and X.W. contributed materials. Z.G., A.G., D.V. and X.W. performed the study. L.Z. provided the analysis code. Z.G. analysed the data. Z.G. and E.J. wrote the original manuscript. All authors edited the manuscript.

Competing financial interests

The authors declare no competing financial interests.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

How concepts are coded in the brain is a core issue in cognitive neuroscience. Studies have focused on how individual concepts are processed, but how conceptual representation changes to suit the context is unclear. We parametrically manipulated the association strength between words, presented in pairs one word at a time using a slow event-related fMRI design. We combined representational similarity analysis and computational linguistics to probe the neurocomputational content of these trials. Individual word meaning was maintained in lateral and medial temporal and medial parietal cortex (areas associated with heteromodal representation) when items were judged to be unrelated, but not when a linking context was retrieved. Context-dependent meaning was represented in left lateral prefrontal and premotor cortex and pre-supplementary motor area (regions associated with semantic control). Moreover, neural representation of context-dependent meaning was affected by the association strength in left lateral prefrontal cortex (LIFG) and anterior temporal lobe (ATL) to different degrees: ATL supported combined meanings more for strong associations; in contrast, LIFG supported combined meanings across trials, even when more control was required. These findings indicate that the semantic store in ATL amplifies long-term semantic associations when these are retrieved but does not directly capture short-term non-dominant associations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The question of how concepts are coded in the brain is a core issue in cognitive neuroscience. Extensive evidence from neuropsychological, neuroimaging and neuromodulation studies has provided information about how individual concepts are represented in the brain (Martin 2007; Patterson et al. 2007; Binder and Desai 2011; Pulvermüller 2013; Yee et al. 2014; Lambon Ralph et al. 2017; Jefferies et al. 2020). However, it is still poorly understood how the brain produces diverse patterns of semantic retrieval to suit the context. For example, APPLE is associated with CAKE when it occurs together with KITCHEN, but also with LAPTOP when we encounter it with KEYBOARD. We therefore create context-dependent representations of concepts, integrating individual meanings into dynamically-modulated multilevel representations to facilitate flexible semantic cognition (Yee and Thompson-Schill 2016).

Even though concepts are thought to be constructed in this dynamic fashion, empirical studies have, until recently, largely focused on invariant conceptual representation – i.e. the features of concepts that do not vary across contexts (Yee and Thompson-Schill 2016). Distributed modality-specific features (e.g. visual, auditory, motor and valence features) in 'spoke' systems are thought to be integrated within a semantic 'hub' or 'convergence zone' in the anterior temporal lobes (ATL), giving rise to heteromodal concepts (Patterson et al. 2007; Lambon Ralph et al. 2017). However, ATL along with left angular gyrus (AG), is also thought to support conceptual combination, with the strongest responses observed when conceptual retrieval is highly coherent and control demands are minimized (Bemis and Pylkkänen 2011, 2013; Teige et al. 2018; Teige, Cornelissen, Mollo, Gonzalez Alam, et al. 2019). In other circumstances, when retrieval must be focused on non-dominant associations and unusual conceptual combinations, there is greater engagement of the 'semantic control network', which includes left inferior frontal gyrus (IFG) and left posterior middle temporal gyrus (pMTG) (Whitney et al. 2010; Hallam et al. 2016; Hallam et al. 2018; Gonzalez Alam et al. 2019; Wang

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

et al. 2020). These semantic control processes can shape the interaction between hub and spokes in order to vary our retrieval of concepts to suit the context in which they occur (Davey et al. 2016; Lambon Ralph et al. 2017; Chiou et al. 2018; Chiou and Lambon Ralph 2019; Zhang et al. 2021).

Multivoxel pattern analysis (MVPA) provides us with a powerful tool to probe how the representation of semantic information in the brain varies according to the context. MVPA studies have started to explore how features combine to form concepts and how word meaning is modified syntactically (Allen et al. 2012; Coutanche and Thompson-Schill 2014; Boylan et al. 2015; Hoffman and Tamm 2020; Solomon and Thompson-Schill 2020). A recent magnetoencephalography study showed that neural representations of the verb DO were modified across temporal, inferior frontal and inferior parietal regions according to the noun it was combined with (Lyu et al. 2019). In these situations, semantic control processes might be able to bias the pattern of semantic retrieval in task-appropriate ways by facilitating or inhibiting connections between the heteromodal hub in ATL and specific task-relevant and task-irrelevant spokes. Yet in many other situations requiring semantic control – for example, when weak as opposed to strong thematic associations must be identified – participants are not required to focus on specific types of features, but instead to identify a context in which concepts co-occur. This might require participants to create an event representation, which can then bias retrieval towards features of the concept that are consistent with this event, and away from other potentially dominant features which are inconsistent (Mirman et al. 2017). An understanding of the neurobiological mechanisms that underpin this process remains elusive.

In the current study, we used fMRI to identify where in the brain non-contextualized meanings of words are represented as well as to determine how words are integrated to form context-dependent conceptual representations. We varied the strength of thematic relationships between two words presented successively, from very strong (dog with leash), through

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

intermediate trials (dog with beach) to very weak pairs (dog with keyboard). We leveraged word embeddings of natural language use to establish vectors of similarity for our word stimuli which were either (i) focused on context-invariant meaning using word2vec (Mikolov et al. 2013) or (ii) captured vectors of similarity for words based on the ongoing context (i.e. taking into account the preceding/following words) using ELMo (Peters et al. 2018). We combined these computational linguistic approaches with a slow-event related fMRI design and representational similarity analysis (RSA) (Kriegeskorte et al. 2006; Kriegeskorte et al. 2008), implemented using a searchlight approach, to determine where in the brain similarity in multi-voxel activity patterns could be predicted by context-free and context-sensitive conceptual similarities.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Materials and Methods

Participants

A group of 32 healthy participants aged 19 to 35 years (mean age = 21.97 ±□3.47 years; 19 females) was recruited from the University of York. They were all right-handed, native English speakers, with normal or corrected-to-normal vision and no history of psychiatric or neurological illness. The study was approved by the Research Ethics Committee of the York Neuroimaging Centre. All volunteers provided informed written consent and received monetary compensation or course credit for their participation. Data from four participants was excluded due to head motion, resulting in a final sample of 28 participants for the semantic task. This study provides a novel analysis of a dataset first reported by (Gao et al. 2020).

Semantic Task

The experimental stimuli were 192 English concrete noun word pairs. We excluded any abstract nouns and pairs of items drawn from the same taxonomic category, so that only thematic links were evaluated. The strength of the thematic link between the items varied parametrically from trials with no clear link to highly related trials; in this way, participants were free to decide based on their own experience if the words had a discernible semantic link. There were no ‘correct’ and ‘incorrect’ responses: instead, we expected slower response times and less convergence across participants for items judged to be ‘related’ when the associative strength between the items was weak, and for items judged to be ‘unrelated’ when the associative strength between the items was strong. Overall, there were roughly equal numbers of ‘related’ and ‘unrelated’ responses across participants.

Each trial began with a visually presented word (WORD-1) which lasted 1.5s, followed by a fixation cross presented at the centre of the screen for 1.5s. Then, the second word (WORD-2) was presented for 1.5s, followed by a blank screen for 1.5s. Participants had 3s from

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the onset of WORD-2 to judge whether this word pair was semantically associated or not by pressing one of two buttons with their right hand (using their index and middle fingers). During the inter-trial interval (3s), a red fixation cross was presented until the next trial began. Both response time (RT) and response choice were recorded. Participants finished 4 runs of the semantic task, each lasting 7.3 min. Before the scan, they completed a practice session to familiarise themselves with the task and key responses.

Neuroimaging Data Acquisition

Imaging data were acquired on a 3.0 T GE HDx Excite Magnetic Resonance Imaging (MRI) scanner using an eight-channel phased array head coil at the York Neuroimaging Centre. A single-shot T2*-weighted gradient-echo, EPI sequence was used for functional imaging acquisition with the following parameters: TR/TE/θ = 1500 ms/15 ms/90°, FOV = 192 × 192 mm, matrix = 64 × 64, and slice thickness = 4 mm. Thirty-two contiguous axial slices, tilted upper to the eye, were obtained to decrease distortion in the anterior temporal lobe and prefrontal cortex. Anatomical MRI was acquired using a T1-weighted, 3D, gradient-echo pulse-sequence (MPRAGE). The parameters for this sequence were as follows: TR/TE/θ = 7.8s/2.3 ms/20°, FOV = 256 × 256 mm, matrix = 256 × 256, and slice thickness = 1 mm. A total of 176 sagittal slices were acquired to provide high-resolution structural images of the whole brain.

Semantic Similarity Matrices

Using natural language processing tools, two semantic similarity matrices were constructed based on two types of word embedding to investigate different types of semantic information in neural activity patterns.

word2vec: The word2vec model represents words as fixed high-dimensional vectors of embeddings. The vectors of word embeddings were generated by training the network on the 100-billion-word Google News corpus. Each time the network was presented with a word from

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the corpus, it was trained to predict the context in which it appeared, where context was defined as the two words preceding and following it in the corpus. The model learns to represent words used in similar contexts with similar patterns; each word's vector had 300 dimensions, with similarity across two words' vectors indicating that they appear in similar contexts, and thus have related meanings. Word2vec embeddings are fixed and unique for each word; for example, irrespective of whether 'apple' was followed/preceded by 'bread' or 'keyboard', its word embeddings were the same. Therefore, using word2vec, we constructed semantic similarity matrices for WORD1 and WORD2 that reflected the meaning of single words/concepts, unmodified by the context in which these items appeared, by calculating cosine similarity between pairs of words drawn from different trials.

ELMo: Given that context can change the meaning of individual words in sentences and phrases, Peters et al. (2018a) proposed a deep contextualized word embedding model called ELMo (Embeddings from Language Models) to capture the context-dependent semantic representation of words. Rather than providing a dictionary of words and their corresponding vectors, ELMo analyses words within their linguistic context, with each token assigned a representation that is a function of the entire input sentence. ELMo representations are deep in the sense that they are a function of all of the internal layers of a deep bidirectional language model: there is a context-independent fixed input vector for the word in the lowest layer, with two higher layers capturing backward and forward context-sensitive aspects of word meaning. We used the pretrained model released by Allennlp (Gardner et al. 2018), which was trained on a large test corpus of 5.5B tokens from Wikipedia and the English news data from the workshop of machine translation (WMT) 2008-2012. We extracted a linear weighted combination across the three layers in ELMo to generate context-sensitive embeddings for WORD2. Each vector representing word meaning had 1024 dimensions. We calculated the current-context-sensitive

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

semantic similarity matrix for WORD-2, to search for brain regions where the pattern of responses across voxels was associated with contextually-constrained semantic cognition.

The association strength for each word-pair measured by ELMo and word2vec was highly similar, the correlation across 192 pairs was strong, $r = 0.706$ ($p < 0.001$). The correlation between the semantic similarity matrix of context dependent meaning for WORD-2 (ELMo) and of original meaning (word2vec) for both two words were examined when all word-pairs were included: WORD-1: $r = -0.018$, $p = 0.0146$; WORD-2: $r = 0.0238$, $p = 0.179$. These weak correlations between ELMo and word2vec based semantic similarity matrices suggested two metrices captured different semantic information.

fMRI Data Preprocessing Analysis

Image preprocessing and statistical analysis were performed using FEAT (FMRI Expert Analysis Tool) version 6.00, part of FSL (FMRIB software library, version 5.0.9, www.fmrib.ox.ac.uk/fsl). The first 4 volumes before the task were discarded to allow for T1 equilibrium. The remaining images were then realigned to correct for head movements. Translational movement parameters never exceeded 1 voxel in any direction for any participant or session. No spatial smoothing was performed. The data were filtered in the temporal domain using a nonlinear high-pass filter with a 100s cutoff. A two-step registration procedure was used whereby EPI images were first registered to the MPRAGE structural image (Jenkinson and Smith 2001). Registration from MPRAGE structural image to standard space was further refined using FNIRT nonlinear registration (Andersson et al. 2007, 2007). The denoised time series were transformed to standard space for the multiple variate analyses.

Pattern Similarity Analysis

In order to examine how the characteristics of semantic representation were influenced by the context, we focused on the decision phase of the task. This period corresponded to TR 6

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

& 7 after WORD-1 onset. Second-order representational similarity analysis (RSA) was performed using a searchlight approach; semantic RSMs (representation similarity matrices, based on semantic similarity) were compared with brain-based RSMs (neural pattern similarity matrices) to test where and what semantic information was represented in the brain. Neural pattern similarity was estimated for cubic regions of interest (ROIs) containing 125 voxels surrounding a central voxel. Within each of these ROIs, we compared patterns of brain activity to derive a neural RSM, derived from the pairwise Pearson correlations of each pair of trials. We excluded any pairs presented in the same run from the calculation of pattern similarity to avoid any auto-correlation issues. Spearman's rank correlation was used to measure the alignment between semantic and brain-based models during the decision phase (the alignment with word2vec and ELMo based semantic model were named as conceptual representation of original meaning and context-sensitive meaning, separately). The resulting coefficients were Fisher's z transformed and statistically inferred across participants. The searchlight analysis was conducted in standard space. A random-effects model was used for group analysis. Since no first-level variance was available, an ordinary least square (OLS) model was used.

Relationship between Neural Pattern Similarity and Semantic Similarity as a Function of Association Strength

To investigate how semantic representation in the brain was modulated by semantic association strength, we performed a novel sliding window pattern similarity analysis. We sorted all the word-pairs from weakly to strongly associated according to their semantic association strength (word2vec value) for the related and unrelated conditions separately. Next, we grouped every 16 trials into one window (window size); adjacent windows partially overlapped with each other (step size) by 4 trials. We then computed second-order RSAs within each window using the above method. The next step of this analysis established how semantic-neural alignment changed as a function of the association strength by using a Spearman correlation. The

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

resulting correlation coefficients were transformed into Fisher's z-scores and then examined using a random-effect model for group analysis, using OLS. We performed several variants of this analysis, using different window and step sizes, in order to ensure the robustness of our conclusions.

Mixed-Effects Modelling Analysis

Since participants judged different numbers of items to be semantically related and unrelated, mixed-effects modelling was used for the analysis of the behavioural and neuroimaging data. This approach is particularly suitable when the number of trials in each condition differs across participants (Mumford and Poldrack 2007; Ward et al. 2013). Mixed-effects modelling was implemented with lme4 in R (Bates et al. 2014). We used the likelihood ratio test (i.e., Chi-Square test) to compare models, in order to determine whether the inclusion of predictor variables significantly improved the model fit.

In the analysis of the behavioural data, semantic association strength was used as a predictor of the decision participants made (judgements of whether the words were related or unrelated) and, in a separate model, the reaction time this decision took. Participant identity was included as a random effect. By comparing models with and without the association strength predictor, we were able to establish whether semantic association strength predicted semantic performance.

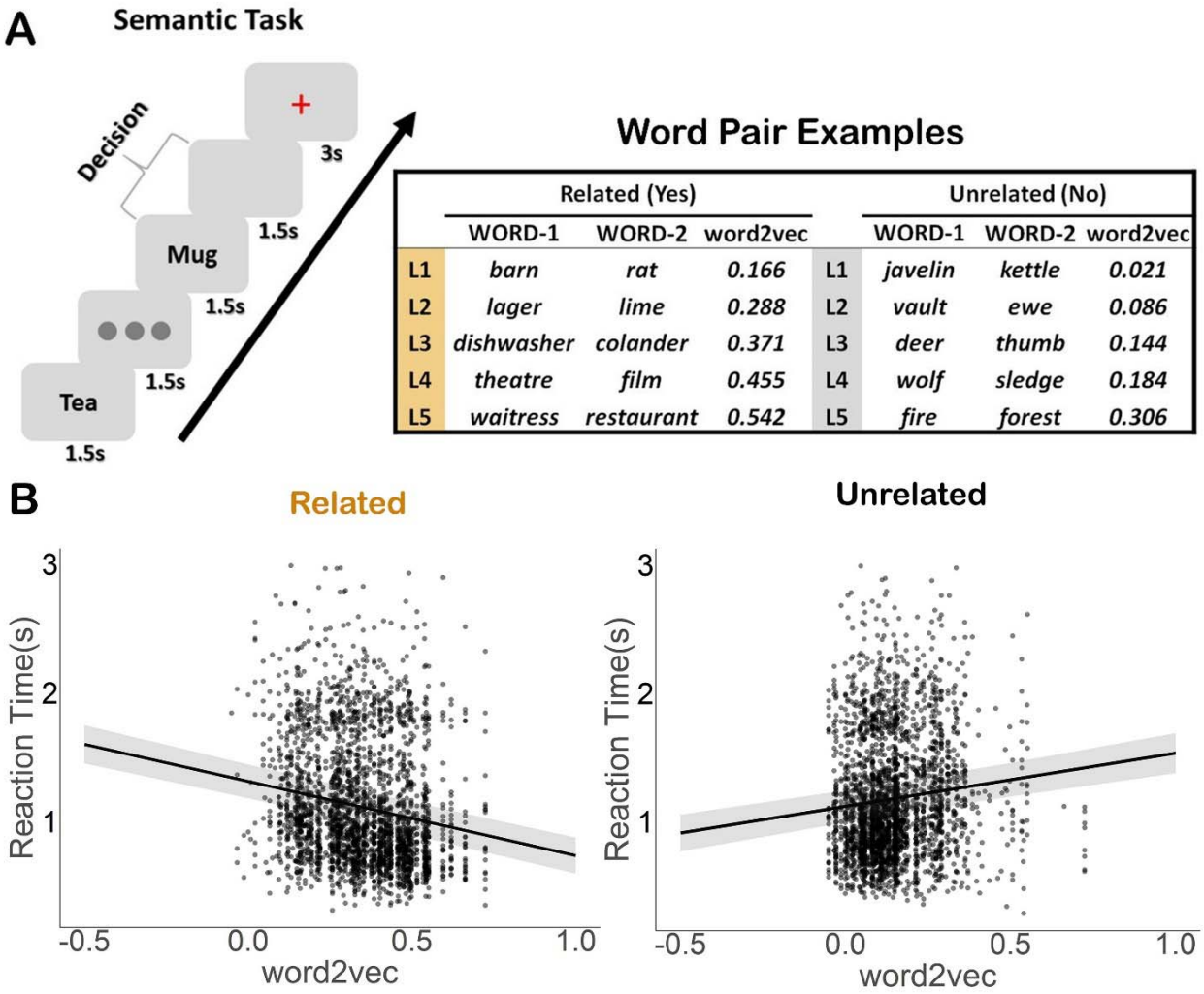
Mixed-effects modelling was also used to examine the influence of association strength on the neural representations of context-dependent meaning (correlation between neural similarity matrix and semantic similarity matrix of ELMo for WORD-2 for each window). Representation of context-dependent meaning was the dependent variable and participant identity was included as a random effect. By comparing this model with and without the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

predictor of association strength, we could uncover whether this association strength could predict conceptual representation in the brain.

Results

Behavioural Results



	Related (Yes)				Unrelated (No)		
	WORD-1	WORD-2	word2vec		WORD-1	WORD-2	word2vec
L1	*barn*	*rat*	*0.166*	L1	*javelin*	*kettle*	*0.021*
L2	*lager*	*lime*	*0.288*	L2	*vault*	*ewe*	*0.086*
L3	*dishwasher*	*colander*	*0.371*	L3	*deer*	*thumb*	*0.144*
L4	*theatre*	*film*	*0.455*	L4	*wolf*	*sledge*	*0.184*
L5	*waitress*	*restaurant*	*0.542*	L5	*fire*	*forest*	*0.306*

Figure 1. Experiment paradigm and behavioural results. A. Left-hand panel: Semantic association task; participants were asked to decide if word pairs were semantically related or not. Right-hand panel: Word pair examples for both related and unrelated decisions from one participant, with association strength increasing from weak (L1; little semantic overlap) to strong

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(L5; high semantic overlap). Trials were assigned to related and unrelated sets on an individual basis for each participant, depending on their decisions, and then split into 5 levels for display purposes, based on word2vec scores. B. The semantic association strength (word2vec) was negatively associated with reaction time for related trials and positively associated with reaction time for trials judged to be unrelated. People were faster to discern a relationship between words when they had high semantic overlap, and slower to decide that the words were unrelated when they had high semantic overlap.

We used linear mixed effects models to examine whether semantic association strength was a reliable predictor of behavioural performance. The strength of the semantic association (word2vec value for each pair) was positively associated with a higher probability that participants would identify a semantic relationship between the words ($\chi2(1) = 2505.4$, $p < 0.001$).

Since we used a continuous manipulation of associative strength, and there is no categorical boundary of word2vec values which can capture the trials reliably judged to be related and unrelated, therefore, traditional error scores were not calculated. Overall, equal numbers of word pairs were judged to be related or unrelated by the participants (mean ratio: 0.491 vs. 0.495, $\chi2(1) = 0.00021$, $p > 0.995$). Linear mixed-effects models examined how association strength modulated reaction time (RT) for trials judged to be related and unrelated. There was a significant effect of strength of semantic association (word2vec) for both related and unrelated decisions: association strength was negatively associated with RT for related trials ($\chi2(1) = 156.55$, $p = 2.2e-16$), and positively associated with reaction time for trials judged to be unrelated ($\chi2(1) = 52.415$, $p =4.5e-13$); see Figure 1B. It was more difficult for participants to retrieve a semantic connection between two words when strength of association was lower;

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

on the contrary, it was easier for them to decide there was no semantic connection between word pairs with low word2vec values.

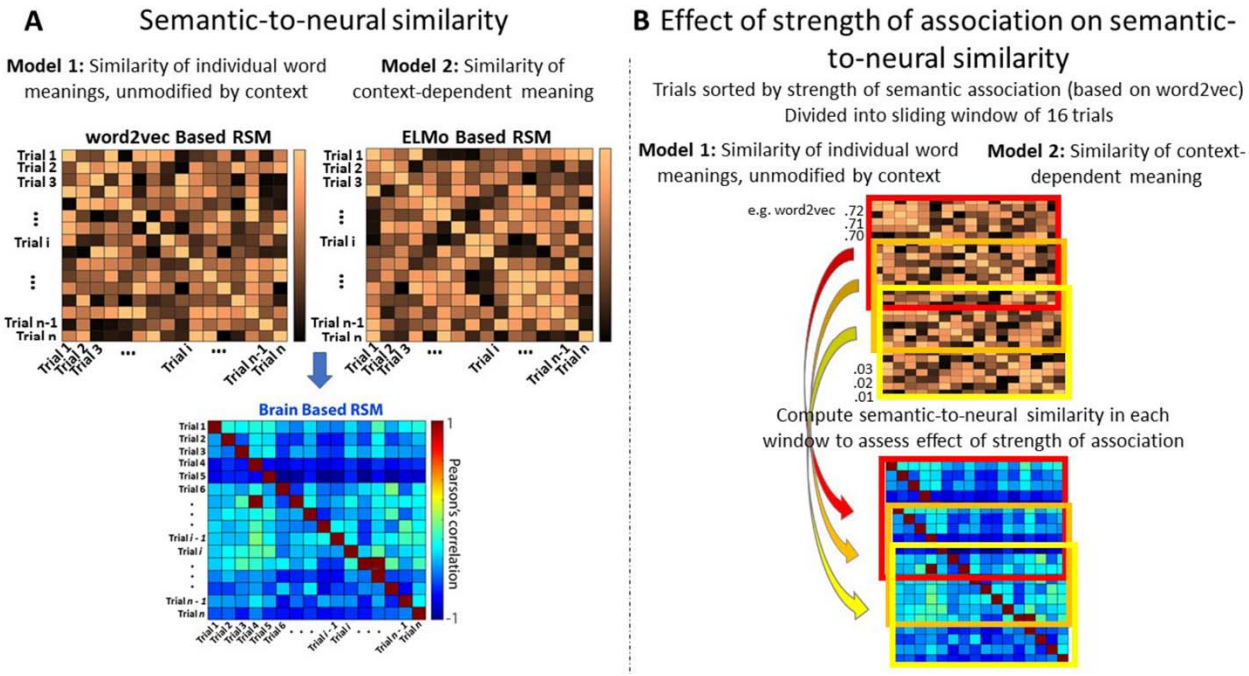


Figure 2. A. Semantic-to-neural similarity computed via second-order RSAs: these analyses characterized the semantic similarity between words on different trials and examined the association with neural similarity across trials. Left-hand panel: word2vec-based RSM for unmodified word meanings across trials – this matrix captured the semantic similarity of individual words used across trials; right-hand panel: ELMo-based RSM for context-dependent meaning – this matrix captured the semantic similarity of contextually-modified meanings across trials. B. Effect of strength of association on semantic-to-neural similarity. Trials were sorted by semantic association strength (based on word2vec-based measures of semantic similarity between word 1 and word 2 within each trial) and were then divided into sliding windows of 16 trials. We computed semantic-to-neural similarity using second-order RSAs in each window, and then correlated association strength with the semantic-to-neural similarity across windows to establish how the representation of context-dependent meaning in the brain changes as a function of the strength of the association between words.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fMRI Results

Neural Representation of Original Meaning (context-free conceptual representation)

Whole-brain analysis was performed using a searchlight approach. First, we examined context-free semantic representation of the original or unmodified meaning, based on the word2vec model, during the decision phase. The strongest representation of context-free meaning is expected for WORD-1, since retrieval of the meaning of this item commenced prior to the semantic context being presented. For trials judged to be semantically unrelated, a significant positive association between neural pattern similarity and semantic similarity based on word2vec was seen across extensive regions of cortex. The neural representation of context-free meaning was strongest in the left anterior inferior and middle temporal gyri, extending to posterior temporal, medial temporal, temporal-occipital and medial parietal cortices; see Figure 3A (left-hand panel). A similar positive relationship between the brain-based model and the word2vec-based semantic model for WORD-2 was observed on trials judged to be semantically unrelated, although this map was more restricted in its extent; see Figure 3A (right-hand panel).

Next, we tested the representation of the original meaning of WORD-1 and WORD-2 for trials judged to be semantically-related. A very different pattern was observed, with only negative correlations between neural and word2vec-based semantic models for both WORD-1 and WORD-2, see Figure 3B. These negative correlations were strongest for WORD-2, when meaning was retrieved in the context of the preceding word, and suggest that the process of identifying a semantic link on related trials pushed the representations of semantically-similar individual items further apart. There was no evidence that the brain maintained the original meaning of single words when participants retrieved a shared meaning-based context for them.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To examine whether the same brain regions supported these distinct effects for trials that participants judged to be related and unrelated, we examined the correspondence between the uncorrected statistical maps reported in Figure 3 (i.e. between the analyses based on word2vec). There were highly significant correlations across all pairs of maps shown in this figure; $p < 2.62e\text{-}80$. The strongest correlation, $r = 0.61$, was between the analysis of WORD-1 for trials judged to be unrelated (Figure 3A, left-hand panel) and of WORD-2 for trials judged to be related (Figure 3B, right-hand panel). This result demonstrates that a common semantic network – including left anterior and posterior lateral temporal regions, medial temporal cortex, angular gyrus, medial parietal cortex and visual regions – supports the representation of original word meaning in the absence of a linking context, and changes this representation when an association is formed between words, such that their meaning is modified.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

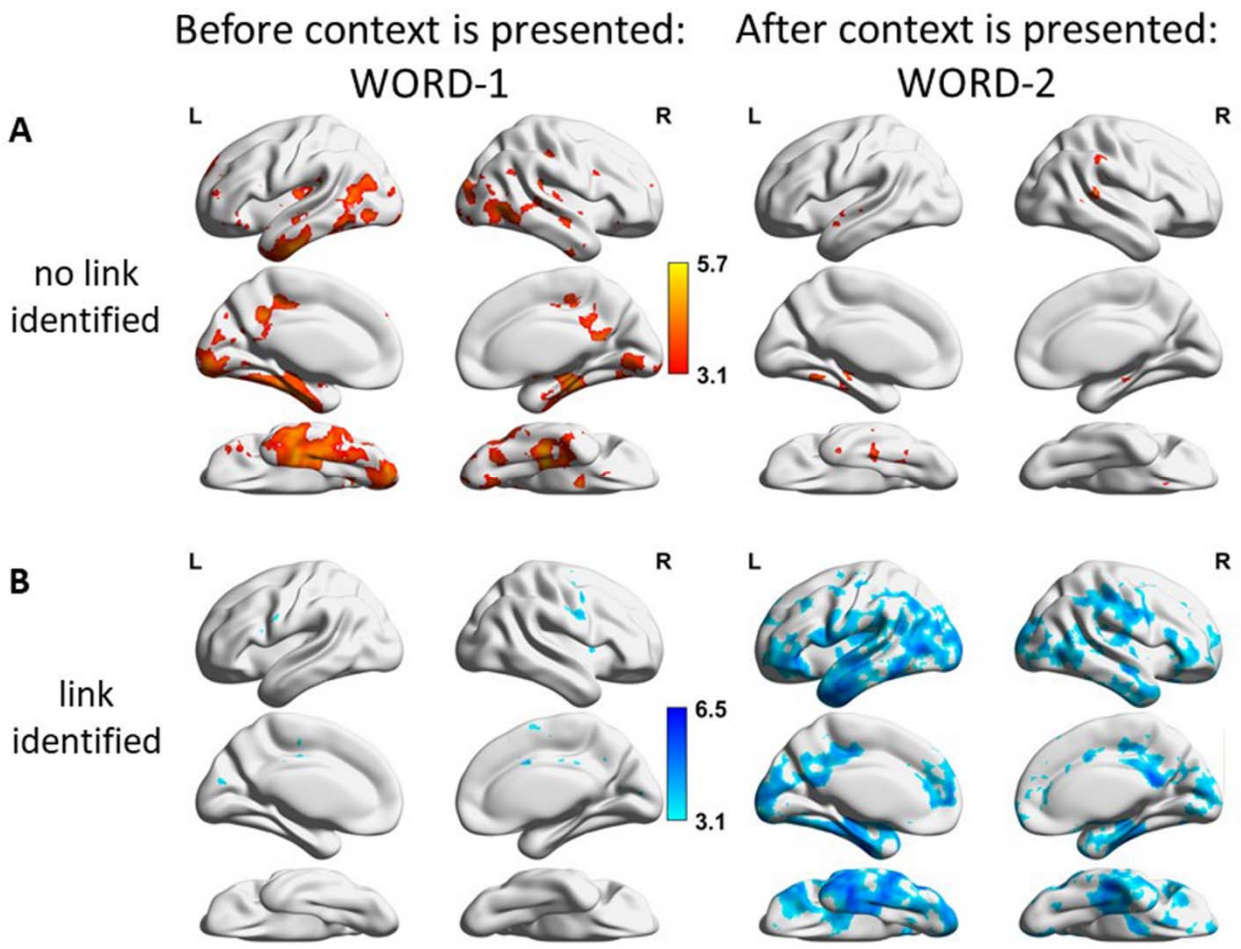


Figure 3. The original meaning of WORD-1 (before context is presented) and WORD-2 (after context is presented) coded in the brain for items judged to be unrelated (A – when no link is identified) and related (B – when a linking context is identified). Positive correlations were found for the unrelated items, while only negative correlations were found for the trials in which people identified a semantic link (Z > 3.1, corrected).

Neural Representation of Context-Dependent Meaning

The preceding results demonstrate that activity patterns in the brain represented the original or unmodified semantic meaning for trials judged to be unrelated, but not for trials in which a linking context was identified. Motivated by the theory that a concept cannot be

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

meaningfully separated from the context in which it occurs (Yee and Thompson-Schill 2016), we next tested whether neural similarity across trials could be related to contextually-derived word meaning, especially for those word pairs judged to be related. We focused this analysis on WORD-2, since the meaning of this item was processed in the context of the preceding item (no semantic context was available when the meaning of WORD-1 was first retrieved). We used ELMo to estimate the context-dependent semantic similarity between the WORD-2 items across trials, separately for words from trials judged to be related and unrelated. For trials judged to be semantically related, a positive correlation between neural similarity and ELMo-based semantic similarity was found in left lateral frontal cortex, precentral gyrus and in pre-supplementary motor area; see Figure 4A (right-hand panel). Only relatively sparse, negative correlations between context-dependent semantic similarity and neural similarity were found for trials judged to be unrelated; see Figure 4A (left-hand panel).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

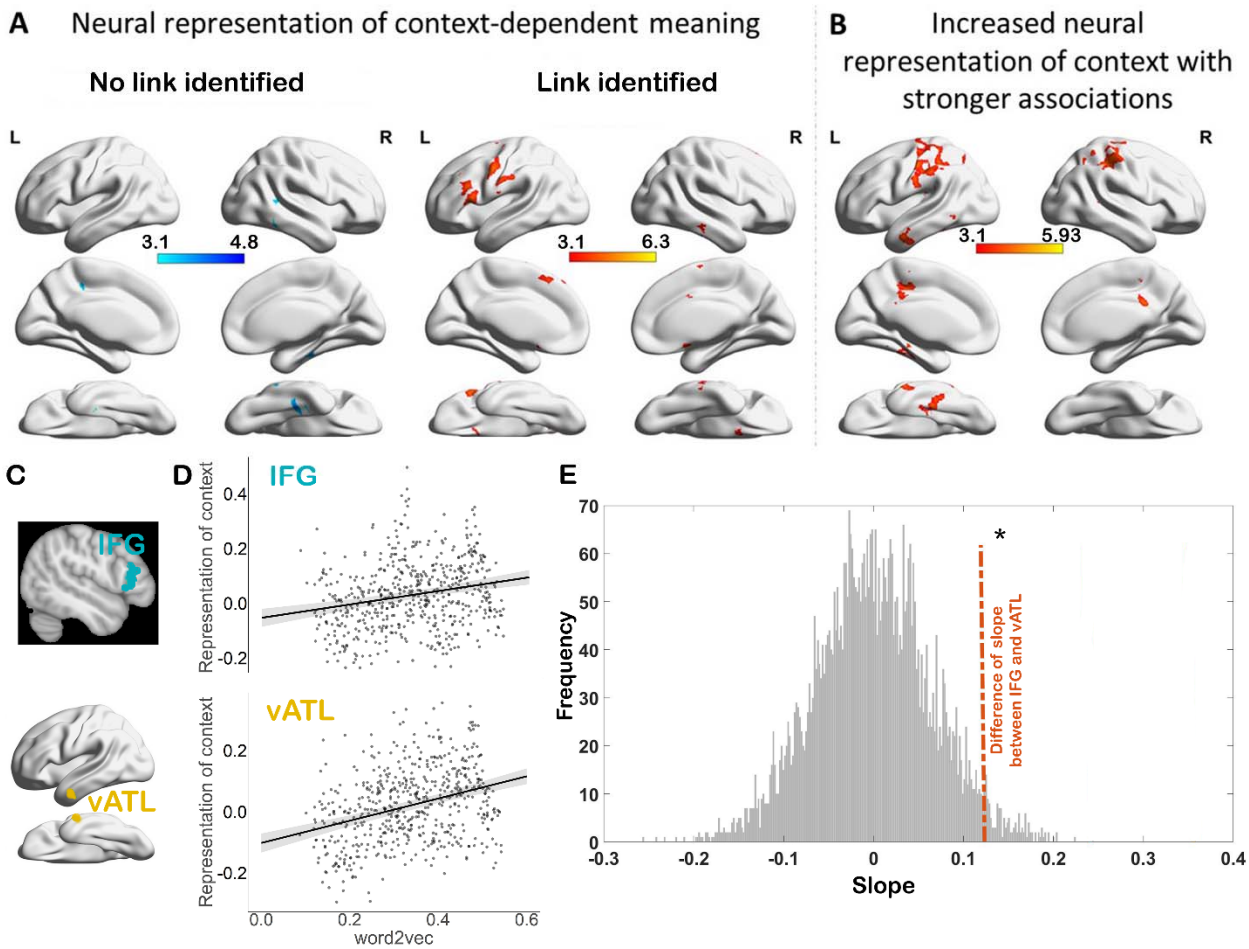


Figure 4. A: The context-dependent meaning of WORD-2 coded in the brain for the trials judged to be unrelated (left-hand panel) and related (right-hand panel) (Z > 3.1, corrected). B: This association between context-sensitive ELMo-based semantic similarity and neural similarity was higher for more strongly-linked word pairs that were judged to be semantically-related (Z > 3.1, corrected). C. Regions of interest in left IFG (aqua) and left ATL (yellow) for which we directly compared the effect of association strength on the representation of contextually-derived meaning. IFG was defined using the cluster representing context-dependent meaning within pars orbitalis and pars triangularis using the Harvard-Oxford brain atlas. Left ATL was defined as the cluster representing both the meaning of individual words in trials judged to be unrelated, and the increased representation of context-dependent meaning for strong associations for related trials, within the anterior temporal cortex using the Harvard-Oxford brain atlas. D. Linear

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mixed-effect modelling found that association strength was more positively associated with neural representation of context-dependent meaning in left ATL compared with IFG. E. Neural representation of contextually-derived meaning showed a significant difference between left ATL and IFG (red dash line). For each window of sixteen trials, we measured the representation of context by calculating the correlation between the neural similarity matrix and corresponding ELMo based similarity, then calculated the mean association strength in each window, we randomly shuffled the neural representation of context and its associated association strength across windows within participants in each ROI, to estimate the slope for the influence of association strength on the neural representation of context-dependent meaning, and to estimate the difference in slopes between ROIs. This procedure was repeated 5000 times to generate the null distribution of the difference between ROIs (gray color).

Context-Dependent Semantic Representation Changed as A Function of Association Strength

Next, we considered how the representation of the context-dependent meaning of WORD-2 changed as a function of association strength. For this analysis, we used RSA combined with a sliding window approach. We sorted related trials according to their association strength (based on a W2V score for each word-pair), from weak to strong associations, and grouped every 16 trials into one window; we then calculated the Spearman's correlation between association strength and the neural representation of meaning. The alignment between neural similarity and context-dependent meaning derived from the ELMo model was greater for more strongly-related trials within left ventral and lateral ATL, and bilateral medial and lateral parietal cortex extending into postcentral gyrus (see Figure 4B). These results indicate that left ATL represents the output of conceptual combinations particularly for strongly associated word-pairs, in line with research suggesting this region is a semantic representational hub that supports the efficient retrieval of

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

strong associations that are well-aligned to the structure of long-term knowledge (Jefferies 2013; Lambon Ralph et al. 2017; Jefferies et al. 2020).

To check the robustness of these results, we generated different window sizes containing different numbers of trials along the continuous dimension of association strength, and changed the extent to which adjacent windows overlapped with each other (i.e. the overlap step size). We confirmed the results were robust across a range of window sizes and overlap step sizes (window sizes and overlapping step sizes of 16,4; 16,8; 20,4, respectively). The positive correlation with association strength for context-dependent meanings in related trials was maintained across these analyses; these results are available at: https://neurovault.org/collections/9399/.

Context-dependent meaning in the left inferior frontal gyrus and anterior temporal lobe was modulated to different degrees by associative strength

According to the controlled semantic cognition framework (Lambon Ralph et al., 2016), when semantic control demands are high – for example, the task requires a semantic link to be identified between weakly-associated items – left IFG is engaged along with other regions of the semantic control network in order to regulate semantic retrieval. In contrast, ventral ATL, as a long-term semantic store, is expected to support the retrieval of strong semantic associations in the absence of additional constraints, since these aspects of knowledge are dominant within this representational system. This account predicts a dissociation between left IFG and ventral ATL in the effect of association strength on the neural representation of context-dependent meanings: ATL should represent the meaning of words in context more strongly when the linking context is strong. IFG is not expected to show this pattern since it is less engaged in these circumstances.

While the results above are consistent with this dissociation between ATL and IFG, they do not permit direct comparison of these sites, because significant whole-brain searchlight results for

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the two sites were identified in different contrasts. To better characterize the similarities and differences in the effect of strength of association on the neural representation of context-dependent meaning, we compared the previously-described effects within two ROIs considering their critical and different roles in semantic processing for weak and strong associations. Left IFG was defined as the cluster representing the context-dependent meaning of WORD-2 for trials judged to be related (Figure 4A, right-hand panel) within pars orbitalis and pars triangularis from the Harvard-Oxford brain atlas, see Figure 4C top-panel. Left ventral ATL was defined using the cluster that represented the meaning of individual words in unrelated trials (Figure 3A, left-hand panel) and the overlapping cluster in which context-dependent meaning was modulated by association strength (Figure 4B), within the anterior division of the temporal lobe from the Harvard-Oxford brain atlas, see Figure 4C bottom panel.

We first examined whether representation of context in the left IFG was modulated by association strength (similar to the pattern in ATL) by extracting the correlation coefficient between association strength and representation of semantic context from the sliding window analysis in the ROI. There was a significant positive correlation in left IFG (Fisher's Z transformed R value, 0.192, $p < 0.0005$). Direct comparison of the correlation coefficients in left IFG and ventral ATL revealed a stronger influence of association strength on neural representations of context-dependent meaning in ATL than IFG (Fisher's Z transformed R value, 0.192 vs 0.327 $p = 0.0392$). The representation of contextually-derived meaning in IFG was modulated by association strength, but this site was less sensitive to association strength than ATL.

We supplemented this analysis with linear mixed-effect modelling including association strength for each trial as the independent variable, participant as a random effect and the representation of context-dependent meaning as the dependent variable. Association strength was positively associated with the representation of the context-dependent meaning of WORD-2 in both left

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

IFG and ventral ATL (slope: b = 0.247, $\chi2(1)$ = 29.185, p =6.5e-8 and b = 0.364, $\chi2(1)$ = 69.129, p =4.4e-16, respectively, Bonferroni corrected), see Figure 4D. Direct comparison of the effect of association strength on the representation of context-dependent meaning in left IFG and ventral ATL (i.e. the slopes for these sites in Figure 4D) showed a smaller decrease in the neural representation of context-dependent meaning from strong to weak associations in left IFG (actual difference between slopes: b = 0.117, p = 0.027, permutation test), see Figure 4E.

The results were highly consistent when spherical ROIs based on MNI coordinates from published studies were used instead of these clusters. We selected a left IFG ROI around the peak voxel reported by (Cardillo et al. 2012; Solomon and Thompson-Schill 2020) at MNI coordinate (x = -54, y = 24, z = 10; 275 voxels). This site has been linked to semantic selection and modulation of conceptual features. We selected a left ventral ATL ROI around the peak reported by (Ueno et al. 2018) at MNI coordinate (x = -47, y = -15, z = -34; also 275 voxels), as this site has been heavily implicated in conceptual combination and representation (Visser et al. 2010; Bemis and Pylkkänen 2011; Visser and Lambon Ralph 2011; Westerlund and Pylkkänen 2014). Linear mixed effect modelling analyses found that the representation of context-dependent meaning was positively associated with association strength in both ROIs (IFG: slope, b = 0.20, $\chi2(1)$ = 19.032, p =1.29e-5; ventral ATL, b = 0.338, $\chi2(1)$ = 53.829, p =2.37e-13, Bonferroni corrected). Compared to left IFG, ventral ATL showed a significantly larger influence of association strength on neural representation of context-dependent meaning (p = 0.018).

Discussion

This study parametrically modulated the association strength between pairs of words to delineate the neural representation of context-free and context-dependent meaning. We used two classes of computational linguistic models, representing concepts as either independent or

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dependent on their linguistic context. Using representational similarity analysis, we found brain activity patterns in the bilateral temporal lobes, medial parietal and left frontal lobes that were associated with context-independent conceptual information – this response was found specifically for trials judged to be semantically unrelated, when there was no linking context to modify the meanings of words. At the same time, context-dependent meanings were represented in regions implicated in semantic control, including left lateral and dorsomedial prefrontal cortex – this response was found specifically for trials judged to thematically related, when a linking context was retrieved. Moreover, neural representation of context-dependent meaning was significantly modulated by association strength: the stronger the association between concepts, the more context-dependent meaning was represented in the left anterior temporal lobe. Finally, the results showed a significant difference between left IFG, implicated in semantic control, and left ventral ATL, implicated in semantic representation, in the impact of associative strength on the representation of context-dependent meaning: the ATL showed stronger effects of association strength, suggesting that this site represents context-dependent meanings for dominant contexts that are encoded strongly in the long-term semantic store.

Past studies have often compared activation patterns elicited by stimuli from different categories, for instance, faces, objects, places and tools; these studies have significantly advanced our understanding of the neural substrates of ‘individual’ (i.e. static) concepts (Binder et al. 2009; Price 2012). Nevertheless, previous behavioural work on conceptual integration has revealed conceptual representation of word meaning is context sensitive; for instance, when ‘red’ is paired with fire, apple or sky, the magnitude of the representation of ‘red’ is modulated by the following noun (Halff et al. 1976; Coutanche et al. 2019). However, the neural mechanisms underlying dynamic conceptual representation in the brain remain largely uncharacterised. Our results address this issue by revealing a shift in the neural representation of semantic information across trials judged to be ‘related’ and ‘unrelated’. For unrelated trials, when the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

meanings of words are retrieved largely independently, much of the semantic network showed a positive correlation between neural similarity and the semantic similarity of individual words estimated from computational linguistic models. These same brain areas showed lower neural similarity with individual meanings when word pairs were judged to be related, because for these trials, the target concept was necessarily processed in the context of the first word. These findings are consistent with studies comparing activation patterns in the brain to semantic models generated from feature norms, behavioural ratings and large text corpora, which have provided converging evidence that the left anterior temporal lobe (ATL) and perirhinal cortex, as well as posterior parietal cortex and left inferior frontal cortex, reflect semantic similarity between concepts (Bruffaerts et al. 2013; Devereux et al. 2013; Clarke and Tyler 2014; Clarke et al. 2015; Borghesani et al. 2016; Carota et al. 2017; Carota et al. 2021). These results validate our adoption of word2vec as a means of capturing context-independent meanings of individual concepts, and also provide clear evidence for the dynamic nature of conceptual representation depending on the context (Willems and Casasanto 2011; Yee and Thompson-Schill 2016).

Previous investigations of dynamic conceptual representation are limited because it is challenging to know how representations of meaning will change between contexts – this information cannot be easily gleaned from participants' reports. ELMo, a recently developed natural language processing algorithm (Peters et al. 2018), allowed us to uncover contextualised conceptual representations in brain regions implicated in semantic control, such as left inferior and middle frontal gyrus, dorsomedial prefrontal cortex and precentral gyrus. These effects were only found when sematic links were identified by participants and not when trials were judged to be unrelated, indicating that semantic control processes play a critical role in semantic integration and/or composition to form coherent meanings (Lambon Ralph et al. 2017). Three recent studies that also employed ELMo to study context-dependent semantic cognition also identified left inferior prefrontal and lateral anterior temporal cortex in context-

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dependent conceptual representation (Lopopolo et al. 2020; Toneva et al. 2020; Lyu et al., 2019) – the current findings add to these observations by revealing functional differences between these frontal and anterior temporal sites in the kinds of context-dependent meanings that they support.

Given our parametric design in which association strength was varied in a continuous fashion, sliding window RSA allowed us to investigate the effect of strength of association on the brain regions that support contextually-guided meaning. With increasing associative strength, we found a larger role for left ventral ATL in the representation of context-dependent meaning. We also found stronger engagement of medial parietal cortex, which might contribute to conceptual representation (Binder and Desai 2011; Krieger-Redwood et al. 2016; Vatansever et al. 2017), and bilateral somatomotor cortex, potentially reflecting the activation of more similar action-based features when context-dependent meanings are related across trials (Liuzzi et al. 2010; Schuil et al. 2013; Carota et al. 2017). This pattern for ATL is consistent with MEG studies finding that strongly associated word-pairs elicit greater neural activity in left ATL (Teige et al. 2018; Teige, Cornelissen, Mollo, Gonzalez Alam, et al. 2019). Our findings are also in line with MEG studies that presented adjective-noun combinations and found stronger engagement of left ATL 200 to 250 ms after noun onset when these words could be combined in meaning (Bemis and Pylkkänen 2013; Pylkkänen 2020). One recent MEG study which monitored participants' brain activity during naturalistic sentence reading found that neural activity patterns in the first 100 to 200ms contained context-independent meaning in the left lateral and anterior temporal lobe, while context-dependent meaning was seen later between 300 ~ 500ms (Lopopolo et al. 2020). Due to the low temporal resolution of fMRI, the current study cannot elucidate the evolution of semantic representation in the brain, but shows differentiation between ATL and left inferior prefrontal cortex in the effect of associative strength; these sites may be harder to separate in M/EEG (Teige, Cornelissen, Mollo, Alam, et al. 2019). We found

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that left inferior frontal cortex but not ATL contributed to the representation of contextually-guided meaning across all the trials. ATL showed a stronger influence of associative strength on the representation of context-derived meaning than IFG, suggesting that it represents contextually-guided meanings when these relate to the dominant contexts in which concepts commonly occur. In this way, ATL only supports semantic contexts that can be readily accessed from long-term knowledge, while IFG additionally sustains more novel contexts that are only weakly encoded in long-term memory. This is consistent with the Controlled Semantic Cognition framework which proposes that these sites provide a heteromodal semantic store (in ATL) and control processes that regulate the retrieval of this information (in IFG) (Jefferies 2013; Lambon Ralph et al. 2017; Hoffman and Tamm 2020; Jackson et al. 2021).

Although studies have looked at how conceptual features are combined in ATL, and how coherent concepts are computed (Coutanche et al. 2019; Pylkkänen 2019), our study provides direct evidence that ATL supports both contextually-independent and contextually-guided meaning retrieval (with the contribution of this site to contextually-guided meaning retrieval restricted to situations in which the linking context is strongly encoded in long-term memory and consequently control demands are minimised). In the past literature, conceptual retrieval was usually constrained by control processes to focus on non-dominant aspects of knowledge when these are required by the task or context. For instance, task requirements can gate the recruitment of ‘spoke’ systems (Zhang et al. 2021); participants can retrieve specific unimodal features when they have task instructions providing a clear goal for conceptual processing, and/or suppress activation of non-relevant spoke representations (Coutanche and Thompson-Schill 2014; Martin et al. 2018). The current study has important theoretical implications since here the task instructions did not change between trials: participants were always judging whether or not the two words were thematically related. In this situation, the meaning of the words themselves defined the nature of the linking context, and established which features

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

should be the focus of subsequent retrieval. This 'stimulus-driven' semantic control appears to be supported by left IFG within the semantic control network, which maintains semantic contexts in a controlled fashion, even when these are non-dominant, to modulate the flow of activation through semantic space. Responses in left ATL did not show the same pattern, suggesting that neural responses within the putative semantic store continued to be governed by dominant aspects of semantic knowledge. The semantic store amplifies long-term semantic associations when relevant but does not directly capture short-term non-dominant associations.

Left IFG has long been linked to semantic selection and control processes (Thompson-Schill et al. 1997; Jefferies 2013; Noonan et al. 2013; Jackson 2020), and is activated during the retrieval of weak semantic associations (Lambon Ralph et al. 2017; Jefferies et al. 2020), but its role in conceptual combination is largely unknown. One recent study found that left IFG is sensitive to feature uncertainty during comprehension of combined concepts, while ATL reflects the integration of conceptual features (Solomon and Thompson-Schill 2020). Another recent study investigated how the brain resolves semantic ambiguity in homonym comprehension and found that IFG supports context-appropriate meaning (Hoffman and Tamm 2020). The current study identifies left IFG as one of a small number of sites that shows increased neural similarity for trials with greater context-dependent neural similarity across all trials – including weak associations when semantic control requirements were higher, in contrast to ATL.

One limitation of the current study was that our measure of context-sensitive conceptual representation (from ELMo) was derived across trials and participants, and was unable to detect individual-specific understanding of each word pair. Moreover, the weaker associations are, the more variance in semantic representation there should be across participants. Future studies could collect subjective reports of context-dependent understanding of word pairs for each participant, and then leverage ELMo to create individual-specific semantic models. More

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

detailed and precise ELMo-based semantic models might result in further neural-semantic alignment results, extending beyond the regions identified here.

In conclusion, this study leverages natural language models and representational similarity analysis, to compare, for the first time, context-independent and context-dependent meaning representation in the brain during sematic decisions. Our study demonstrates that overlapping brain regions support context-independent and context-dependent meaning but that the semantic network also shows differences in the effect of associative strength between words. Left IFG represents the current meaning-based context needed to support ongoing semantic cognition across all trials, including when this context is only weakly represented in long-term memory, and consequently control processes may be needed to regulate retrieval. In contrast, ventral ATL amplifies strongly-encoded long-term semantic associations when these are relevant to ongoing cognition, but does not directly capture short-term non-dominant associations. These findings clarify the roles of semantic control regions and the putative ATL semantic hub in the computation of coherent meanings across inputs.

References

Allen K, Pereira F, Botvinick M, Goldberg AE. 2012. Distinguishing grammatical constructions with fMRI pattern analysis. Brain and Language. 123:174-182.

Andersson JL, Jenkinson M, Smith S. 2007. Non-linear registration aka Spatial normalisation FMRIB Technial Report TR07JA2. FMRIB Analysis Group of the University of Oxford.

Andersson JL, Jenkinson M, Smith S. 2007. Non-linear registration, aka spatial normalisation. FMRIB technial report TR07JA2. 22.

Bates D, Mächler M, Bolker B, Walker S. 2014. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:14065823.

Bemis DK, Pylkkänen L. 2011. Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. Journal of Neuroscience. 31:2801-2814.

Bemis DK, Pylkkänen L. 2011. Simple Composition: A Magnetoencephalography Investigation into the Comprehension of Minimal Linguistic Phrases. The Journal of Neuroscience. 31:2801-2814.

Bemis DK, Pylkkänen L. 2013. Flexible composition: MEG evidence for the deployment of basic combinatorial linguistic mechanisms in response to task demands. PloS one. 8:e73949.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Binder JR, Desai RH. 2011. The neurobiology of semantic memory. Trends in cognitive sciences. 15:527-536.

Binder JR, Desai RH, Graves WW, Conant LL. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral cortex. 19:2767-2796.

Borghesani V, Pedregosa F, Buiatti M, Amadon A, Eger E, Piazza M. 2016. Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding. NeuroImage. 143:128-140.

Boylan C, Trueswell JC, Thompson-Schill SL. 2015. Compositionality and the angular gyrus: A multi-voxel similarity analysis of the semantic composition of nouns and verbs. Neuropsychologia. 78:130-141.

Bruffaerts R, De Deyne S, Meersmans K, Liuzzi AG, Storms G, Vandenberghe R. 2019. Redefining the resolution of semantic knowledge in the brain: Advances made by the introduction of models of semantics in neuroimaging. Neuroscience & Biobehavioral Reviews. 103:3-13.

Bruffaerts R, Dupont P, Peeters R, De Deyne S, Storms G, Vandenberghe R. 2013. Similarity of fMRI Activity Patterns in Left Perirhinal Cortex Reflects Semantic Similarity between Words. The Journal of Neuroscience. 33:18597-18607.

Cardillo ER, Watson CE, Schmidt GL, Kranjec A, Chatterjee A. 2012. From novel to familiar: tuning the brain for metaphors. Neuroimage. 59:3212-3221.

Carota F, Kriegeskorte N, Nili H, Pulvermüller F. 2017. Representational Similarity Mapping of Distributional Semantics in Left Inferior Frontal, Middle Temporal, and Motor Cortex. Cerebral Cortex. 27:294-309.

Carota F, Nili H, Pulvermüller F, Kriegeskorte N. 2021. Distinct fronto-temporal substrates of distributional and taxonomic similarity among words: evidence from RSA of BOLD signals. NeuroImage. 224:117408.

Chiou R, Humphreys GF, Jung J, Lambon Ralph MA. 2018. Controlled semantic cognition relies upon dynamic and flexible interactions between the executive 'semantic control' and hub-and-spoke 'semantic representation' systems. Cortex. 103:100-116.

Chiou R, Lambon Ralph MA. 2019. Unveiling the dynamic interplay between the hub- and spoke-components of the brain's semantic system and its impact on human behaviour. NeuroImage. 199:114-126.

Clarke A, Devereux BJ, Randall B, Tyler LK. 2015. Predicting the time course of individual objects with MEG. Cerebral Cortex. 25:3602-3612.

Clarke A, Tyler LK. 2014. Object-specific semantic coding in human perirhinal cortex. Journal of Neuroscience. 34:4766-4775.

Coutanche MN, Solomon S, Thompson-Schill SL. 2019. Conceptual Combination in the Cognitive Neurosciences.

Coutanche MN, Thompson-Schill SL. 2014. Creating Concepts from Converging Features in Human Cortex. Cerebral Cortex. 25:2584-2593.

Davey J, Thompson HE, Hallam G, Karapanagiotidis T, Murphy C, De Caso I, Krieger-Redwood K, Bernhardt BC, Smallwood J, Jefferies E. 2016. Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. Neuroimage. 137:165-177.

Devereux BJ, Clarke A, Marouchos A, Tyler LK. 2013. Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. Journal of Neuroscience. 33:18906-18916.

Gao Z, Zheng L, Chiou R, Gouws A, Krieger-Redwood K, Wang X, Varga D, Ralph ML, Smallwood J, Jefferies E. 2020. Distinct and Common Neural Coding of Semantic and Non-semantic Control Demands. bioRxiv.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gardner M, Grus J, Neumann M, Tafjord O, Dasigi P, Liu N, Peters M, Schmitz M, Zettlemoyer L. 2018. Allennlp: A deep semantic natural language processing platform. arXiv preprint arXiv:180307640.

Gonzalez Alam TRdJ, Karapanagiotidis T, Smallwood J, Jefferies E. 2019. Degrees of lateralisation in semantic cognition: Evidence from intrinsic connectivity. NeuroImage. 202:116089.

Halff HM, Ortony A, Anderson RC. 1976. A context-sensitive representation of word meanings. Memory & Cognition. 4:378-383.

Hallam GP, Thompson HE, Hymers M, Millman RE, Rodd JM, Lambon Ralph MA, Smallwood J, Jefferies E. 2018. Task-based and resting-state fMRI reveal compensatory network changes following damage to left inferior frontal gyrus. Cortex. 99:150-165.

Hallam GP, Whitney C, Hymers M, Gouws AD, Jefferies E. 2016. Charting the effects of TMS with fMRI: Modulation of cortical recruitment within the distributed network supporting semantic control. Neuropsychologia. 93:40-52.

Hoffman P, Tamm A. 2020. Barking up the right tree: Univariate and multivariate fMRI analyses of homonym comprehension. NeuroImage. 219:117050.

Jackson RL. 2020. The Neural Correlates of Semantic Control Revisited. bioRxiv.2020.2007.2015.204990.

Jackson RL, Rogers TT, Lambon Ralph MA. 2021. Reverse-engineering the cortical architecture for controlled semantic cognition. Nature Human Behaviour.

Jefferies E. 2013. The neural basis of semantic cognition: converging evidence from neuropsychology, neuroimaging and TMS. Cortex. 49:611-625.

Jefferies E, Thompson H, Cornelissen P, Smallwood J. 2020. The neurocognitive basis of knowledge about object identity and events: dissociations reflect opposing effects of semantic coherence and control. Philosophical Transactions of the Royal Society B. 375:20190300.

Jenkinson M, Smith S. 2001. A global optimisation method for robust affine registration of brain images. Medical image analysis. 5:143-156.

Krieger-Redwood K, Jefferies E, Karapanagiotidis T, Seymour R, Nunes A, Ang JWA, Majernikova V, Mollo G, Smallwood J. 2016. Down but not out in posterior cingulate cortex: Deactivation yet functional coupling with prefrontal cortex during demanding semantic cognition. NeuroImage. 141:366-377.

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. Proceedings of the National Academy of Sciences. 103:3863-3868.

Kriegeskorte N, Mur M, Bandettini PA. 2008. Representational similarity analysis-connecting the branches of systems neuroscience. Frontiers in systems neuroscience. 2:4.

Lambon Ralph MA, Jefferies E, Patterson K, Rogers TT. 2017. The neural and computational bases of semantic cognition. Nature Reviews Neuroscience. 18:42.

Liuzzi G, Freundlieb N, Ridder V, Hoppe J, Heise K, Zimerman M, Dobel C, Enriquez-Geppert S, Gerloff C, Zwitserlood P, Hummel FC. 2010. The Involvement of the Left Motor Cortex in Learning of a Novel Action Word Lexicon. Current Biology. 20:1745-1751.

Lopopolo A, Schoffelen JM, van den Bosch A, Willems RM. 2020. Words in context: tracking context-processing during language comprehension using computational language models and MEG. bioRxiv.2020.2006.2019.161190.

Lyu B, Choi HS, Marslen-Wilson WD, Clarke A, Randall B, Tyler LK. 2019. Neural dynamics of semantic composition. Proceedings of the National Academy of Sciences. 116:21318-21327.

Martin A. 2007. The representation of object concepts in the brain. Annu Rev Psychol. 58:25-45.

Martin CB, Douglas D, Newsome RN, Man LL, Barense MD. 2018. Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. eLife. 7.

Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J editors. Distributed representations of words and phrases and their compositionality, Advances in neural information processing systems; 2013. 3111-3119 p.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mirman D, Landrigan J-F, Britt AE. 2017. Taxonomic and thematic semantic systems. Psychological bulletin. 143:499.

Mumford JA, Poldrack RA. 2007. Modeling group fMRI data. Social cognitive and affective neuroscience. 2:251-257.

Noonan KA, Jefferies E, Visser M, Lambon Ralph MA. 2013. Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. Journal of cognitive neuroscience. 25:1824-1850.

Patterson K, Nestor PJ, Rogers TT. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. Nature reviews neuroscience. 8:976-987.

Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L. 2018. Deep contextualized word representations. arXiv preprint arXiv:180205365.

Price CJ. 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage. 62:816-847.

Pulvermüller F. 2013. How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. Trends in cognitive sciences. 17:458-470.

Pylkkänen L. 2019. The neural basis of combinatory syntax and semantics. Science. 366:62-66.

Pylkkänen L. 2020. Neural basis of basic composition: what we have learned from the red–boat studies and their extensions. Philosophical Transactions of the Royal Society B: Biological Sciences. 375:20190299.

Schuil K, Smits M, Zwaan R. 2013. Sentential Context Modulates the Involvement of the Motor Cortex in Action Language Processing: An fMRI Study. Frontiers in Human Neuroscience. 7.

Solomon SH, Thompson-Schill SL. 2020. Feature Uncertainty Predicts Behavioral and Neural Responses to Combined Concepts. The Journal of Neuroscience. 40:4900-4912.

Teige C, Cornelissen PL, Mollo G, Alam TRdJG, McCarty K, Smallwood J, Jefferies E. 2019. Dissociations in semantic cognition: Oscillatory evidence for opposing effects of semantic control and type of semantic relation in anterior and posterior temporal cortex. cortex. 120:308-325.

Teige C, Cornelissen PL, Mollo G, Gonzalez Alam TRDJ, Mccarty K, Smallwood J, Jefferies E. 2019. Dissociations in semantic cognition: Oscillatory evidence for opposing effects of semantic control and type of semantic relation in anterior and posterior temporal cortex. Cortex. 120:308-325.

Teige C, Mollo G, Millman R, Savill N, Smallwood J, Cornelissen PL, Jefferies E. 2018. Dynamic semantic cognition: Characterising coherent and controlled conceptual retrieval through time using magnetoencephalography and chronometric transcranial magnetic stimulation. Cortex. 103:329-349.

Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. 1997. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proceedings of the National Academy of Sciences. 94:14792-14797.

Toneva M, Mitchell TM, Wehbe L. 2020. The meaning that emerges from combining words is robustly localizable in space but not in time. bioRxiv.2020.2009.2028.316935.

Ueno T, Meteyard L, Hoffman P, Murayama K. 2018. The Ventral Anterior Temporal Lobe has a Necessary Role in Exception Word Reading. Cerebral Cortex. 28:3035-3045.

Vatansever D, Bzdok D, Wang H-T, Mollo G, Sormaz M, Murphy C, Karapanagiotidis T, Smallwood J, Jefferies E. 2017. Varieties of semantic cognition revealed through simultaneous decomposition of intrinsic brain connectivity and behaviour. Neuroimage. 158:1-11.

Visser M, Jefferies E, Lambon Ralph M. 2010. Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. Journal of cognitive neuroscience. 22:1083-1094.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.03.442424; this version posted May 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Visser M, Lambon Ralph M. 2011. Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. Journal of cognitive neuroscience. 23:3121-3131.

Wang X, Margulies DS, Smallwood J, Jefferies E. 2020. A gradient from long-term memory to novel cognition: Transitions through default mode and executive cortex. NeuroImage. 220:117074.

Ward EJ, Chun MM, Kuhl BA. 2013. Repetition suppression and multi-voxel pattern similarity differentially track implicit and explicit visual memory. Journal of Neuroscience. 33:14749-14757.

Westerlund M, Pylkkänen L. 2014. The role of the left anterior temporal lobe in semantic composition vs. semantic memory. Neuropsychologia. 57:59-70.

Whitney C, Kirk M, O'Sullivan J, Lambon Ralph MA, Jefferies E. 2010. The Neural Organization of Semantic Control: TMS Evidence for a Distributed Network in Left Inferior Frontal and Posterior Middle Temporal Gyrus. Cerebral Cortex. 21:1066-1075.

Willems RM, Casasanto D. 2011. Flexibility in embodied language understanding. Frontiers in psychology. 2:116.

Xu Y, Wang X, Wang X, Men W, Gao J-H, Bi Y. 2018. Doctor, Teacher, and Stethoscope: Neural Representation of Different Types of Semantic Relations. The Journal of Neuroscience. 38:3303-3317.

Yee E, Chrysikou EG, Thompson-Schill SL. 2014. Semantic memory.

Yee E, Thompson-Schill SL. 2016. Putting concepts into context. Psychonomic bulletin & review. 23:1015-1027.

Zhang M, Varga D, Wang X, Krieger-Redwood K, Gouws A, Smallwood J, Jefferies E. 2021. Knowing what you need to know in advance: The neural processes underpinning flexible semantic retrieval of thematic and taxonomic relations. NeuroImage. 224:117405.