A common framework for semantic memory and semantic composition

Ryan M.C. Law*, Olaf Hauk, Matthew A. Lambon Ralph

MRC Cognition & Brain Sciences Unit, University of Cambridge, Cambridge, United Kingdom

*For correspondence: ryan.law@mrc-cbu.cam.ac.uk

Number of pages: 25
Number of figures: 6
Number of tables: 3
Number of words: 252 (Abstract), 684 (Introduction), 1563 (Discussion)

Abstract

How the brain constructs meaning from individual words and phrases is a fundamental question for research in semantic cognition, language and their disorders. These two aspects of meaning are traditionally studied separately, resulting in two large, multi-method literatures, which we sought to bring together in this study. Not only would this address basic cognitive questions of how semantic cognition operates but also because, despite their distinct focuses, both literatures ascribe a critical role to the anterior temporal lobe (ATL) in each aspect of semantics. Given these considerations, we explored the notion that these systems rely on common underlying computational principles when activating conceptual semantic representations via single words, vs. building a coherent semantic representation across sequences of words. The present study used magnetoencephalography and electroencephalography to track brain activity in participants reading nouns and adjective-noun phrases, whilst integrating conceptual variables from both literatures: the concreteness of nouns (e.g., "lettuce" vs. "fiction") and the denotational semantics of adjectives (subsective vs. privative; e.g., "bad" vs. "fake"). Region-of-interest analyses show that the left, and to a lesser extent the right, ATLs responded more strongly to phrases at different timepoints, irrespective of concreteness, suggesting a unified ATL function for semantic memory and composition. Decoding analyses further revealed a time-varying representational format for adjective semantics, whereas representations of noun concreteness were more stable and maintained for around 300 ms. Further, the neural representation of noun concreteness was modulated by the preceding adjectives: decoders learning concreteness signals in single words generalized better to subsective relative to privative phrases.

Keywords: semantic cognition, semantic memory, semantic composition, magnetoencephalography, electroencephalography, anterior temporal lobe

1. Introduction

Constructing meaning from individual words and sequences of words are fundamental challenges for research in semantic cognition, language and their disorders. These two aspects of semantics have formed separate research traditions, namely semantic memory and semantic composition. In the present study, we sought to unify these two literatures under a common framework. Not only would this effort address fundamental questions of how semantic cognition works, but also because, despite their distinct focuses, both literatures consistently implicate the anterior temporal lobe (ATL). Motivated by these two principal reasons, we explore the idea that activating conceptual semantic representations via single words, vs. building up meaning across sequences of words, might share underlying neurocomputational principles.

One perspective on the neurocomputational bases of semantic memory is the hub-and-spokes (H&S) model (for reviews, see e.g., Rogers et al., 2004; Patterson et al., 2007; Lambon Ralph et al., 2017). ATL involvement in semantic memory is evidenced most notably in the syndrome of semantic dementia, in which atrophy to the ATLs leads to progressive deterioration in verbal and non-verbal conceptual knowledge (Patterson et al., 2007). On the other hand, semantic composition, or conceptual combination, has been investigated extensively in a minimal two-word phrase reading paradigm (for reviews, see e.g., Coutanche et al., 2019; Pylkkänen, 2019, 2020; Călinescu et al., 2023). These two large literatures have a striking convergence: they both implicate the left, and to a lesser extent the right, ATLs. Does the overlap in anatomical and functional localization hint at shared computations across the two systems?

A recent computational model by Hoffman et al. (2018) offers a framework that converges models of semantic memory and semantic composition. The model takes a sequence of single words as input and learns by predicting both the upcoming words and each word's sensory-motor properties (for the concrete words alone). Through this learning process, the model acquires context-sensitive semantic representations that reflect the multimodal experiential knowledge for each concept (as described in the standard H&S model) while also accounting for co-occurrence and information integration over time (McClelland et al., 1989; St. John & McClelland, 1990; Rabovsky et al., 2018). Indeed, because non-verbal experiences unfold over time, for concepts to be coherent and generalizable, this knowledge must be integrated both across modalities and across time (Jackson et al., 2021; Lambon Ralph et al., 2010). By assimilating evidence from these different strands, we propose that the ATLs provide a unified function essential to semantic representation and thus play a central role in both parallel literatures.

If there is indeed a unified transmodal, transtemporal integrated semantic "hub" function for the ATL, then adjective-noun processing should be sensitive to variables that are associated with individual concept strength which have been extensively studied in healthy and patient populations for single concept explorations. One of the most prominent of these variables is concreteness or imageability. To our knowledge, however, investigations into adjective-noun processing have primarily focused on nouns denoting concrete concepts. Thus, we varied the concreteness of the nouns in adjective-noun pairs (e.g., *bad lettuce*/*fiction*). We also took the opportunity to integrate a different kind of semantic distinction of *denotation*, introduced by subsective vs. privative adjectives (e.g., *bad* vs. *fake*; see Methods), which significantly modulates interpretation (Kamp, 1975; Partee, 2010; Del Pinal, 2015). For example, "this is a bad lettuce" implies "this is a lettuce", whereas "this is a fake lettuce" implies "this is not a lettuce"—a shift in category membership. Similar contrasts have been explored recently, and has been found to modulate late ERP responses at around 500 ms onwards (Fritz & Baggio, 2020).

If our notion is correct, then we would expect adjective-noun combinatorial meaning to be influenced by these conceptual variables. In a pre-registered magnetoencephalography (MEG) and electroencephalography (EEG) study, we charted neural activity as 36 volunteers silently read nouns and adjective-noun phrases. Using mass-univariate region-of-interest analyses, we tested (1) whether neural systems supporting composition are generalized engines, supporting the composition of concrete and abstract concepts alike, and (2) whether subtle adjective semantic distinctions are reflected within ATL signals, suggesting sensitivity to phrasal interpretation. Using multivariate decoding methods, we also tracked the representational dynamics of compositional semantics.

2. Methods

The study was approved by the local ethics committee. The study was pre-registered on the Open Science Framework on July 14, 2023, before data collection began. The pre-registration document can be accessed at https://osf.io/m82nb/.

2.1 Participants

36 healthy adult participants (27 women; mean age = 26 years; SD = 6 years, range 18-40 years) participated in the study. The number of participants were calculated based on a power estimation (effect size $f = 0.25$, type I error rate $\alpha = 0.01$, type II error rate $\beta = 0.05$) using *G*Power* (v3.1.9.6; Faul et al., 2007) on effect sizes reported in Bemis and Pylkkänen (2011). Participants were native English speakers, right-handed, without history of neurological disorder, had normal or corrected-to-normal vision, and were able to undergo magnetic resonance imaging (MRI).

Participants gave informed consent and were paid £12/hour for MEG and EEG, £12/hour for anatomical MRI scans, and £6/hour for stimulus norming (web-based sample separate from the MEG-EEG sample).

2.2 Stimuli

We chose to manipulate the concreteness of nouns because concrete and abstract nouns have been argued to have distinct representational formats (Paivio, 1971; Schwanenflugel & Shoben, 1983). As we are interested in integrating the semantic memory and semantic composition literatures, we imported this conceptual variable into a phrasal composition paradigm (see Procedure below) and explored how the variables interact.

For adjectives, we drew on insights from linguistic theory and Natural Language Processing. Our main semantic contrast here is one of denotational semantics, associated with subsective and privative adjectives. Subsective adjectives (e.g., *blue*, *new*) function to create a subset of the denotations of the noun. Consider the expression "*blue diamond*". Its denotation is a subset of the denotations of "*diamond*", in that the statement "This is a blue diamond" entails "this is a diamond". In contrast, privative adjectives (e.g., *fake*, *false*) behave differently with regards to entailment: "This is a fake diamond" does not entail "this is a diamond". One way to conceptualize this distinction is that privative adjectives change the category membership of the phrase in a way that subsective adjectives do not. This contrast was chosen as they may reveal how the conceptual representation of concreteness is modulated by the powerful interpretational effects driven by the adjectives, allowing us to further probe interactions between semantic memory and composition.

To create our stimuli, we selected nouns spanning a broad range of conceptual domains while controlling for key psycholinguistic properties. Using concreteness ratings from Brysbaert et al. (2014), we sampled words denoting concrete and abstract concepts. The norm contains human concreteness ratings range from 1 (fully abstract) to 5 (fully concrete). Concrete words were drawn from the range of 4.0-5.0, while abstract words were drawn from the range of 1.5-2.5; These ranges reflect the two local maxima in the distribution of mean concreteness ratings. From this pool, we used *LexOPS* (Taylor et al., 2020) to create concrete-abstract word pairs matched on variables known to influence lexical and conceptual processing, such as unigram Zipf frequency, age of acquisition, and valence (Table 1). All selected words were known by at least 90% of the raters, 3-10 letter long, noun-classified, and had low rating variability (SD < 1). We manually excluded items that were overly evocative or inappropriate, resulting in 137 matched pairs. These words span diverse conceptual categories, including objects, animals, food, sounds, places, emotions, and situations.

Table 1. Psycholinguistic characteristics of the final set of 100 nouns in each condition.

	Concrete	Abstract
Concreteness rating	4.55 (0.29)	2.09 (0.27)
Imageability	5.96 (0.66)	3.28 (0.59)
Length (number of letters)	5.80 (1.32)	6.06 (1.24)
Zipf frequency	4.11 (0.68)	4.14 (0.67)
Age of acquisition (years)	4.60 (2.82)	5.49 (2.85)
Valence	5.00 (1.79)	5.08 (2.03)
Lexical decision latency (ms)	570.02 (51.24)	569.27 (46.58)

Mean values for each measure are reported, with SD in parentheses.

We drew on insights from the field of Natural Language Processing to inform our selection of adjectives. Privative adjectives were sourced from the list compiled by Nayak and colleagues (2014), while subsective adjectives were taken from Lalisse and Asudeh's (2015) compilation. From each category, we selected 16 adjectives chosen for their broad compatibility with our noun set. To create the single-word condition, we generated length-matched consonant strings that served as visual placeholders lacking semantic content.

To create phrases, each noun was randomly paired with an adjective. As a coarse proxy of naturalness, we used Google Books Ngram data (Michel et al., 2011) to compute bigram frequencies for each phrase by summing the bigram frequencies within a set. Based on these aggregate scores, we selected the top 100 stimulus sets (600 items total; 100 per condition) for our final stimulus list. The number of items per condition adopted here exceeds prior MEG studies that used ~40 items per condition (Bemis & Pylkkänen, 2011; Pylkkänen, 2019, 2020), to better power our study to detect potentially subtle effects of denotation and concreteness. Table 2 shows an example set; For the full stimuli list, see https://osf.io/m82nb/.

To encourage participants to process the stimuli at the conceptual level, we opted for a phrase comprehension task (e.g. the stimulus "*bad lettuce*" followed by "*Is it crunchy?*") instead of a picture-matching task as used in the semantic memory literature (e.g., Rogers et al., 2006) and the sentence processing literature (e.g., Bemis & Pylkkänen, 2011). This decision was based on two considerations: First, a picture-matching task is not suitable for abstract items. Second, following our manipulation of denotational semantics (subsective vs. privative), a comprehension task ensures that we probe compositional semantic processing, rather than simply aspects of individual adjectives or nouns. We created comprehension probes for 10% of trials (n=60). We note that the probe questions were designed to engage comprehension but were not matched for length or normed for difficulty. Consequently, the behavioral results reported here are largely descriptive and were not critical to the conclusions we drew.

Table 2. An illustrative set of items showing the study design.

"Adjective" type	Noun type	Stimulus 1	Stimulus 2
Consonant string	Concrete	xtq	lettuce
Subsective adjective	Concrete	bad	lettuce
Privative adjective	Concrete	fake	lettuce
Consonant string	Abstract	xtq	reason
Subsective adjective	Abstract	bad	reason
Privative adjective	Abstract	fake	reason

2.3 Experiment procedure

Items were fully randomized and divided into five blocks (a brief break every 8 minutes or so). The stimuli appeared in light grey against black, backprojected onto a monitor about 90 cm away from the participant's eyes. Each word or letter string subtended maximally at a horizontal visual angle of 4° and a vertical visual angle of .6°. Each trial began with a fixation cross in the center of the screen for 300 ms, followed immediately by stimulus 1 (either letter strings or adjectives) on for 300 ms and off for 300 ms, stimulus 2 (noun) on for 300 ms and off for 500 ms. The decision to allow for a longer period post second word onset (800 ms) is based on Fritz and Baggio's (2020) finding that denotation effects emerge late (500-800 ms). After 10% of trials there was a yes/no comprehension question, presented in whole and in cyan, prompting a right-hand button press response: index finger for yes and middle finger for no. A further inter-trial interval of 1.5 s on average (uniformly jittered between 0.75-1.25 s) was included for cortical activations to return to baseline (Hansen et al., 2010) (Figure 1). The whole experiment inside the MEG chamber took around 45 minutes (excluding EEG preparation time, instructions, etc.).

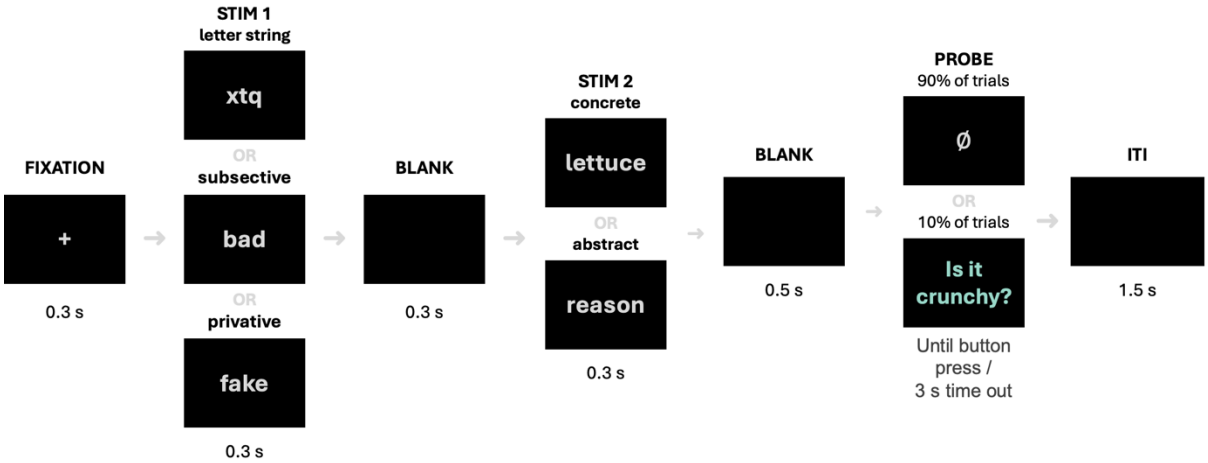


Figure 1. Trial structure. Task probes were included for 10% of trials to encourage attention.

2.4 MEG-EEG data acquisition and preprocessing

We collected concurrent MEG and EEG recordings to maximize spatial resolution (Molins et al., 2008) as well as structural MR images for each participant. MEG and EEG data were acquired in a magnetically shielded room using a 306-channel TRIUX neo system (MEGIN, Espoo, Finland), with 204 planar gradiometers and 102 magnetometers. EEG data were collected concurrently using a 64-electrode EEG cap (EasyCap GmbH, Herrsching, Germany) with a 10-20 layout. EEG reference and ground electrodes were attached to the nose and left cheek, respectively. The electrooculogram (EOG) was recorded by electrodes above and below the left eye (vertical EOG) and at the outer canthi (horizontal EOG), respectively. Impedance levels were reduced to 10 kOhms or lower during preparation when possible. Prior to data acquisition, the positions of 5 head position indicator coils attached to the EEG cap, 3 anatomical landmark points (two preauricular points and nasion) as well as approximately 50–100 additional points covering the scalp were digitized using a 3Space Isotrak II System (Polhemus, Colchester, Vermont, USA) for later co-registration with MRI data. EEG and MEG data were digitized at a sampling rate of 1 kHz and with a high-pass filter of 0.03Hz. As for MR images, for each participant, a T1-weighted structural scan was acquired using a Siemens 3T Prisma scanner.

We then used the Python package *MNE-Python* (v1.4.2; Gramfort et al., 2013, 2014) to pre-process the MEG and EEG data. Data were bandpass-filtered offline between 0.1-40 Hz with a finite impulse response filter design using a Hamming window method. Data were subject to the following noise reduction procedures. For MEG data, we first identified and repaired bad and noisy channels. We then applied Signal Space Separation (SSS) with its spatiotemporal extensions to reduce environmental artefacts (Taulu & Kajola, 2005). This step also included head movement compensation. For EEG, bad channels were visually identified and interpolated using the spherical spline method (Perrin et al., 1989). To attenuate eye-movement artefacts, we performed an independent component analysis (ICA) using the *FastICA* algorithm (Hyvärinen & Oja, 2000). Eye-movement related components were detected using correlation with EOG channels and were inspected visually and removed. Filtered and noise-reduced continuous data were then segmented into epochs from 0.2 s before fixation cross onset to 1.4 s post first word onset. Baseline correction was applied using signals spanning 0.2 before fixation cross onset. We applied a common baseline for all epochs. Epoched data were then downsampled to 250 Hz.

Epochs were first rejected using an automated procedure using the Python package *autoreject* (Jas et al., 2017). All trials were included in further analysis since the comprehension trials are to encourage attention and comprehension. This procedure yielded on average around 80 trials per condition for further analyses.

2.5 Source reconstruction

We used *FreeSurfer* (v6.0.0; Fischl, 2012) alongside *MNE-Python* (v1.4.2; Gramfort et al., 2013, 2014) to reconstruct subject-specific anatomical models of the scalp, inner/outer skull, and cortical surface. T1-weighted anatomical MR images were first segmented using the automated segmentation algorithms in *FreeSurfer* (Dale et al., 1999) to obtain the reconstructed scalp surface.

Forward and inverse solutions were calculated using *MNE-Python*. We used a three-layer boundary element (BEM) model to compute the forward model of each participant's scalp, outer skull surface, and inner skull surface (conductivity parameters are 0.3, 0.006, and 0.3 respectively). The MEG-EEG sensor configurations and MR structural images were then co-registered by matching the scalp digitization points to the scalp surface. Subject-specific source spaces contained 4098 dipole sources in each hemisphere. To calculate the inverse solution, we estimated the noise covariance matrix using the 200 ms window prior to fixation and then regularized it using an automated model selection method (Engemann & Gramfort, 2015). The regularized noise covariance matrices and forward solutions were used to assemble the inverse operator using L2 minimum norm estimation (L2-MNE) with a loose orientation constraint value of 0.2 and depth weighting of 0.8. The inverse operator was then regularized assuming a signal-to-noise ratio (SNR) of 3. After applying the inverse operator to sensor evoked responses, individual-specific cortex-wide source estimates were morphed to a common space ('*fsaverage'*) for group analyses.

2.6 Analyses

2.6.1 Mass-univariate region-of-interest (ROI) analyses

Our regions of interest are the ATL, temporo-parietal cortex (TPC), posterior temporal lobe (PTL), and inferior frontal cortex (IFC). The ATL has been implicated in conceptual processing in studies of semantic memory (e.g., Rogers et al., 2006; Patterson et al., 2007; Lambon Ralph et al., 2017), while the ATL and TPC has been implicated in conceptual processing in phrase and sentence processing studies (Matchin et al., 2019; Pallier et al., 2011; Pylkkänen, 2019, 2020). The IFC and PTL have been implicated in studies of semantic control (Jackson et al., 2021, 2021; Lambon Ralph et al., 2017).

For our ROI definition, we selected four MNI coordinates (see Table 3Table 3) on the cortical surface that correspond to the ROIs, and created a 30-mm (40-mm for the ATLs) label extent around each (Figure 2). These ROI labels encapsulate peaks and clusters found in previous phrase comprehension studies using MEG (Bemis & Pylkkänen, 2011; Pylkkänen, 2019, 2020) and functional MRI (Pallier et al., 2011) and semantic cognition studies on and meta-analyses of fMRI studies (Lambon Ralph et al., 2017; Jackson, 2021). In addition, we also included their right hemisphere homologues by reversing the sign of the x-coordinate to create labels.

Table 3. MNI coordinates of ROIs.

ROIs	MNI coordinates	ROIs	MNI coordinates
Left ATL	[-57, -2, -33]	Right ATL	[57, -2, -33]
Left PTL	[-50, -52, -6]	Right PTL	[50, -52, -6]
Left IFC	[-45, -66, 24]	Right IFC	[45, -66, 24]
Left TPC	[-51, 30, 6]	Right TPC	[51, 30, 6]

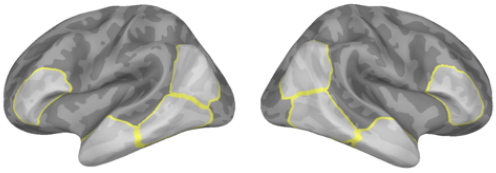


Figure 2. Regions of interest.

We performed separate non-parametric cluster-based permutation tests (Maris & Oostenveld, 2007) in each ROI time-series data using the Python package *Eelbrain* (version 0.39; Brodbeck et al., 2023). First, source-localized cortical activity within each ROI was averaged across sources, yielding one average time course per ROI. Depending on the research questions, we fitted different statistical models to the ROI data at each time sample separately in the whole 800 ms of the second-word window for each ROI separately. In each analysis, subjects were included as random effects. To test for composition effects, we fitted 2x2 ANOVAs, including fixed-effects factors concreteness (concrete, abstract) and structure (single word control, subsective phrase). To test for denotation effects, we fitted 2x2 ANOVAs, including fixed-effects factors concreteness (concrete, abstract) and denotation (subsective, privative). Cluster-level p values were estimated via Monte Carlo simulations, repeated 5,000 times. Cognizant of the inferential limits of cluster-based permutation methods, we follow the suggestions of Sassenhagen and Draschkow (2019) when reporting clusters.

2.6.2 Multivariate decoding analyses

This analysis was exploratory and thus not pre-registered. Temporal decoders attempt to find a linear combination of MEG and EEG sensors that reliably predicts the experimental condition of a given trial (e.g., whether a given trial is concrete or abstract trial). For each participant, decoders were evaluated at the time point they were trained, resulting in a time course of decoding performance summarized as an area under the receiver operating characteristic curve (AUC) score. These time courses indicate whether and when neural signals contain information about a given condition of interest. Temporal generalization (King & Dehaene, 2014), in addition, asks whether a decoder trained at time t is able to generalize to all other time instants. The result is a train-time by test-time matrix summarizing whether neural representations are stable in time. Thus, the diagonal of this matrix is the decoding performance at each time point, as described before. In sum, diagonal and off-diagonal decoding reveals the content encoded

within neural signals, while off-diagonal performance provides additional information about the stability of these representations over time.

Input features to the decoding models were sensor data from all available MEG and EEG sensors. To improve the SNR of the input data, we created "pseudotrials" by averaging several trials together, depending on the specific contrasts we were interested in: For subsective/privative adjective semantics, we averaged pairs of trials across the two levels of concreteness (80 trials each) within each denotation level, resulting in around 80 trials per adjective type. For concrete/abstract noun semantics, we averaged three trials across the three levels of adjective types (80 trials each) within each concreteness level, resulting in around 80 trials per noun type. We then performed temporal decoding analyses, training and testing a separate decoder at each time point to distinguish between concrete vs. abstract trials or subsective vs. privative trials.

To address the question of how adjective semantics shape noun representations, we trained decoders to distinguish concrete vs. abstract in single words and tested their ability to generalize separately to subsective and privative phrasal contexts. For this, we also similarly performed averaging separately within each test condition, yielding approximately 40 trials per condition. Similar to above, we performed a temporal decoding analysis, fitting a decoder at each time point to distinguish between concrete vs. abstract trials. This analysis was similarly exploratory and thus not pre-registered. However, recent EEG work exploring similar contrasts in univariate ERP responses informed our analysis (Fritz & Baggio, 2020). The authors explored denotational aspects of meaning and identified a cluster around 500-800 ms post noun onset. We expected, in our analysis, that the impact of our adjective manipulation would result in differential decodeability similarly later, but not early, in neural activity. We took the cluster extent that Fritz and Baggio (2020) found to define a "late" time window, spanning 500-800 ms). As for an "early" time window, we drew insights from the well-characterized N400 ERP component and its relevance in meaning processing (Kutas & Federmeier, 2011) and defined a 200 ms window from 300-500 ms (See Figure 6). To statistically test for interactions between phrase type (subsective vs. privative) and time window (early vs. late), we fitted a linear mixed effects model to averaged neural responses within each time window using the *lme4* package in R (Bates, Mächler, Bolker, & Walker, 2015). The model included, as fixed effects: phrase type (subsective, privative), time window (early, late), an interaction term (phrase type by time window). Random effects include a by-participant random intercept and a by-participant random slope for phrase type (more complex random effects structure did not converge).

Data from each sensor were standardized separately using RobustScaler from *scikit-learn* (Pedregosa et al., 2011). The Scaler removes the median and scales the data using its interquartile range, as this method is more robust to outliers compared to the approach of removing the mean and scaling to unit variance. All decoding analyses were performed within-subjects and within a stratified ten-fold cross validation scheme. Decoders were L2-regularized logistic regression models implemented using *scikit-learn* with default parameters (C=1). For each subject, decoding performance was summarized with an area under the curve (AUC) score for each decoder for each fold then averaged across fold. For group-level statistics for both diagonal and off-diagonal decoding, the subject-level scores were then submitted to cluster-based one-tailed permutation tests against a chance level of 0.5.

2.6.3 Behavioral analyses

As aforementioned, comprehension probes engaged comprehension and attention but were not matched for length or normed for difficulty, so behavioral data were reported descriptively and were not central to our conclusions.

To analyze reaction time data, we excluded incorrect trials. We fitted a linear mixed effects regression model to group-level data using the *lme4* package in R (Bates, Mächler, Bolker, &

Walker, 2015). The model included fixed effects for concreteness (concrete, abstract), denotation (baseline, subsective, privative), and an interaction term (concreteness by denotation). To analyze response accuracy, we fitted a generalized mixed effects logistic regression to trial responses (this time incorrect responses included). For random effects, both models included by-participant and by-item varying intercepts (more complex random effects structure did not converge). Estimated marginal means were estimated using the R package *emmeans* (Lenth, 2024). ANOVAs were conducted using the R package *car* (Fox & Weisberg, 2019).

3. Results

3.1 Behavioral results

For accuracy, we found a main effect of denotation (χ^2=20.10, p<.0001) and an interaction between denotation and concreteness (χ^2=7.86, p<.05). There was no main effect of concreteness (χ^2=3.18, p>.05). Participants were most accurate at answer comprehension probes for concrete subsective phrases (e.g., *bad lettuce*; M=92%, SE=3%), followed by abstract subsective phrases (e.g., *fake lettuce*; M=74%, SE=6%), single word baselines (e.g., *lettuce/fiction*; M=74-80%, SE=5-6%) then privative phrases (e.g., *fake lettuce/fiction*; M=56-63%, SE=7-8%).

As for reaction times, we found main effects for concreteness (χ^2=19.38, p<.0001) and denotation (χ^2=15.50, p<.0005) but no interactions (χ^2=2.14, p>.05). Participants were faster overall for concrete (M=1.66 s, 95% CI [1.53, 1.80]) relative to abstract items (M=1.92 s, 95% CI [1.78, 2.06]), in line with reaction time findings using other behavioral tasks in previous reports (Hoffman, 2016). In terms of denotation, reaction times parametrically varied as a function of levels of denotation. Subsective phrases were fastest (M=1.64 s, 95% CI [1.49, 1.79]), followed by single-word baselines (M=1.82 s, 95% CI [1.67, 1.97]), followed by privative phrases (M=1.91 s, 95% CI [1.77, 2.06]). Shorter reactions for subsective relative to single word baselines replicated previously reported "phrasal advantage" (Bemis & Pylkkänen, 2011; Rabagliati et al., 2017). In addition, longer reaction times for privative phrases reflect additional processing costs associated with privative semantics, in line with predictions from linguistic theory (Partee, 2010) and processing models (Coulson & Fauconnier, 1999).

3.2 The ATLs as generalized engines for semantic composition

Across all tested ROIs, we found reliable main effects of composition (p = .0008; p = .027) in the left ATL, and a marginal effect of composition (p = .085) in the right ATL, regardless of concreteness (Figure 3). In the left ATL, clusters corresponding to this effect were localized at around 390-660 ms post-noun-onset, with subsective phrases eliciting higher activity relative to single-word controls (cluster 1 Cohen's f = 0.12; cluster 2 Cohen's f = 0.10). In the right ATL, the effect corresponded to a cluster around 420-450 ms post-noun-onset where source activity was, again, higher for phrases compared to single words (Cohen's f = 0.12). See Supplemental Figure 1 for the time courses of all tested ROIs and Supplemental Figure 2 for uncorrected spatiotemporal t-maps.

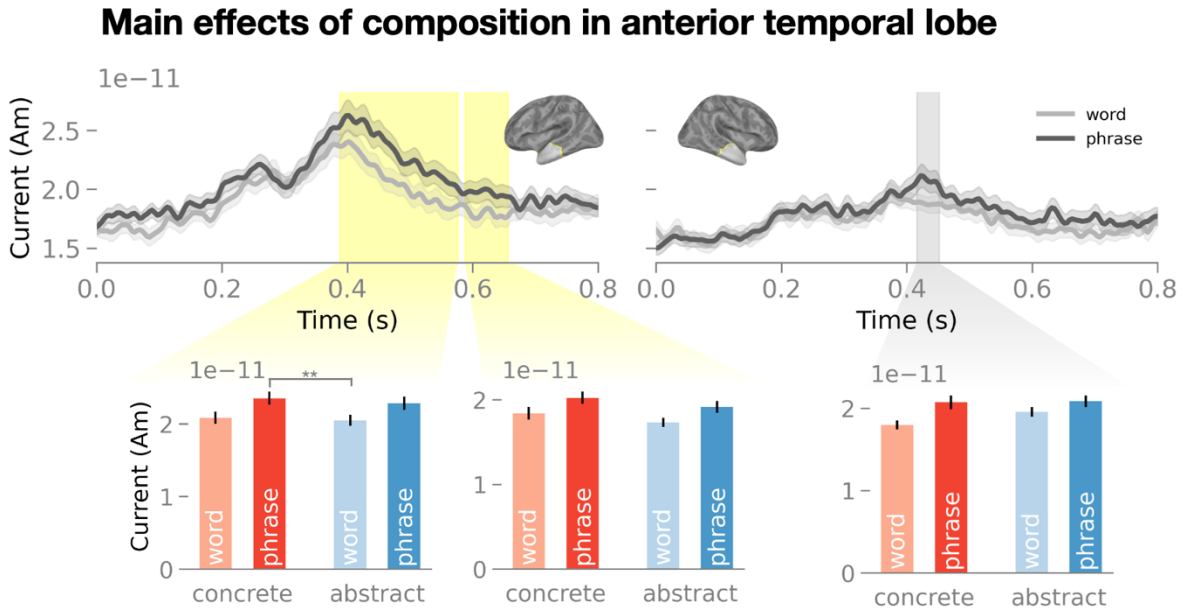


*Figure 3. Main effects of composition in the left and right anterior temporal lobes (ATLs), respectively. Time series plots depict ATL activity elicited by words and subsective phrases. Time zero is noun onset. Significant clusters are shaded in yellow and marginally significant clusters are shaded in grey. Error bars represent one within-subjects SEM (Loftus & Masson, 1994). Bar plots expand the averaged activity by condition in each cluster. **: p < .01.*

3.3 ATL signals show sensitivity to adjectival types associated with phrasal interpretation

We found a marginal effect of denotation in the left ATL (p = .052) but not in other tested ROIs. The associated cluster spanned 380-440 ms post-noun-onset (Figure 4), with increased activity for subsective relative to privative phrases (Cohen's f = 0.08). Like the main effects of composition above, concreteness did not modulate effects of denotation. See Supplemental Figure 3 for the time courses of all tested ROIs and Supplemental Figure 4 for uncorrected spatiotemporal t-maps.

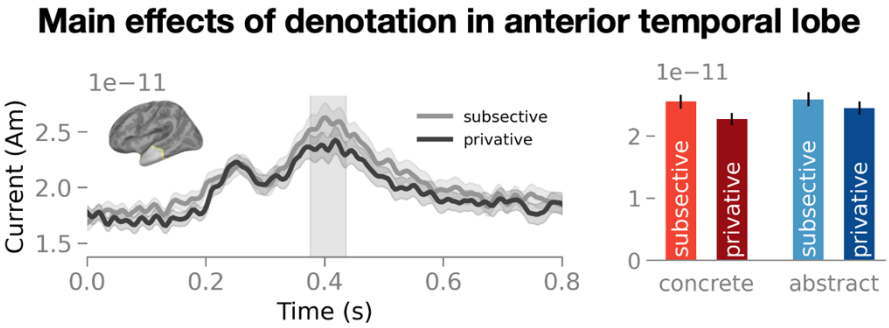


Figure 4. Main effect of denotation in the left anterior temporal lobe (ATL). Time series plot depicts ATL activity elicited by subsective and privative phrases. Time zero is noun onset. Significant clusters were shaded in grey. Error bars represent one within-subjects SEM (Loftus & Masson, 1994). Bar plots expand the averaged activity by condition in the cluster.

3.4 Distinct representational formats for adjective and noun semantics

To characterize how individual adjective and noun semantics are represented, we trained decoders on each time-point separately and tested those decoders on held-out data at those time points. Adjective semantics (subsective vs. privative) were decodable around 320-730 ms post adjective onset (p = .0002; Figure 5A), while noun semantics (concrete vs. abstract) were decodable 400-600 ms post noun onset (p = .0002, p = .0484; Figure 5C).

We then used temporal generalization to test whether the representations of adjective and noun semantics are maintained in a stable format during phrase comprehension. Cluster-based permutation tests revealed above-chance off-diagonal decoding for both adjective (p = .005; Figure 5B) and noun semantics (p = .004; Figure 5D). Diagonal decoding patterns revealed that the format of the representation of adjective semantics is dynamically evolving over time. This implies that the brain encodes adjective semantics at time points between around 300 and 700 ms post adjective onset but the neural patterns changed from time point to time point. In contrast, the format of the representation of noun semantics appeared to be more stable across time, characterized by off-diagonal decodeability. For example, decoders trained at 1.0 s were able to generalize up to 1.3 s and vice versa, implying a recurring, stable format for representing conceptual semantics between approximately 400-700 ms post noun set.

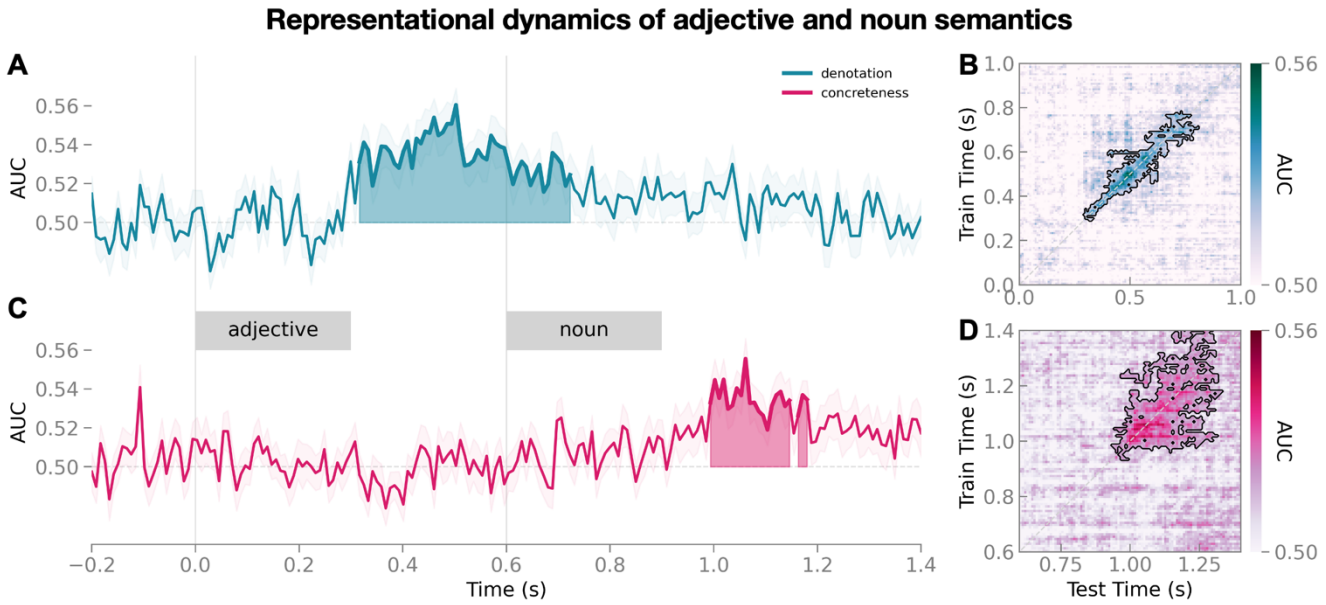


Figure 5. Decoding adjective and noun semantics. (A) Diagonal decoding time course and (B) Temporal generalization matrix (TGM) of denotational semantics (subsective vs. privative), where time 0 is adjective onset. (C) Diagonal decoding time course and (D) TGM of conceptual semantics (concrete vs. abstract), where time 0.6 is noun onset. AUC: Area under the curve summarizing decoding performance. Error bars represent one within-subjects SEM (Loftus & Masson, 1994). Colored areas under decoding time series and black outline in TG matrices denote significant clusters corresponding to group-level effects assessed with cluster-based one-sample permutation tests against chance.

3.5 Adjective semantics differential impact noun representation

After characterizing the temporal representation of these concepts, we sought to understand how noun representations evolve as a function of adjective semantics. In this analysis, decoders were trained on single word trials, distinguishing concrete vs. abstract. The decoders were then evaluated on how well they generalized to subsective and on privative trials separately. Cluster-based permutation tests on the entire decoding score time series revealed above chance decoding in subsective phrases (p = .002; Figure 6A) and not in privative phrases (Figure 6C). The cluster extended from 370 to 450 ms post noun onset. We followed up with linear mixed effects modelling to explore the interactions between phrase type (subsective vs. privative) and time window (early vs. late) and found a main effect of time window (R^2 = .46, p = .037; Figure 6E) and no interaction. While the analysis did not reveal a statistically significant interaction, it is worth noting the graded nature of decodeability as a function of adjective type and time window (Figure 6E).

To test the stability of concreteness representation across time, we again performed a temporal generalization analysis with cluster-based permutation tests, which revealed above-chance off-diagonal decoding in subsective phrases (p = .020; Figure 6B) but not in privative phrases (Figure 6D). Off-diagonal decodeability suggests that conceptual representations of concreteness are maintained in subsective phrases, implying a relatively stable representational format.

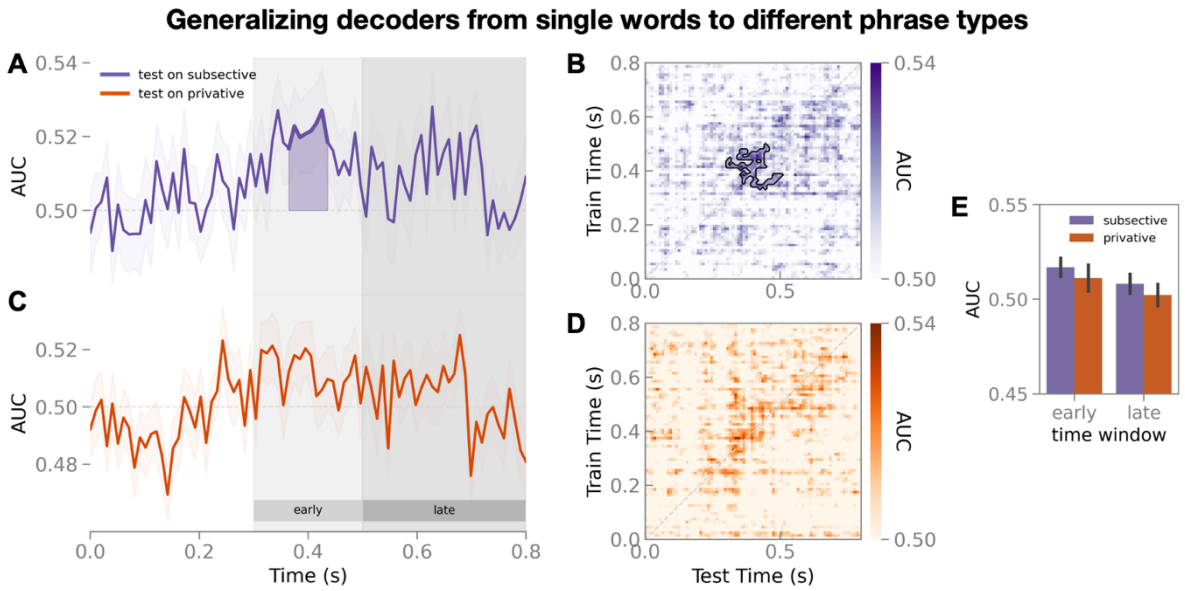


Figure 6. Decoders trained on single word concreteness generalized to subsective but not privative phrases. (A, C) Diagonal decoding performance over time, where time 0 is noun onset. (B, D) Temporal generalization matrices, where train time corresponds to single words and test time corresponds to either subsective phrases (B) or privative phrases (D), where time 0 is noun onset. Error bars represent one within-subjects SEM (Loftus & Masson, 1994). Colored areas under decoding time series denote clusters corresponding to significant group-level effects assessed with cluster-based one-sample permutation tests against chance. (E) Bar plot showing averaged responses by time window and test type.

4. Discussion

A key challenge in research on semantic cognition, language and their impairments is how meaning is constructed at the level of individual words as well as phrases. These two dimensions of semantics have traditionally been studied separately, within the domains of semantic memory and semantic composition. In integrating these research areas, the present study aimed not only to address fundamental questions about how semantic cognition operates but was also motivated by the observation that both literatures, despite their distinct focuses, consistently highlight the role of the anterior temporal lobe (ATL) in conceptual processing. In an experimental framework, we investigated whether the activation of conceptual semantic representations from single words and the construction of coherent meaning across word sequences rely on similar neurocomputational mechanisms.

Our analyses revealed that the ATLs are implicated in and sensitive to both concrete and abstract combinatorial semantics across diverse adjective-noun combinations. The left, and to a lesser extent the right, ATL responded more strongly to phrases than words, independent of concreteness (Figure 3). The left ATL was also sensitive to whether a noun was preceded by a subsective vs. privative adjective (Figure 4). In addition, multivariate decoding analyses, which provided a complementary means in characterizing mechanistic details of conceptual processing, revealed that the representational dynamics of adjective and noun semantics were distinct: adjective representations were time-varying, while noun representations were more stable in time (Figure 5). By using a cross-decoding regime, we show that adjectives shapes noun semantic representations, with concreteness signals in single words being in a more consistent format in subsective phrases than privative phrases (Figure 6).

In short, the ATLs are engaged in the processing of concrete and abstract composition and are also sensitive to phrasal level interpretation. Taken together, the present results are compatible with the idea that the ATL provides a unified function for semantic memory and semantic composition.

How might the two aspects of semantic cognition and language be unified within a common framework? Computational modelling efforts have provided formal, "open science" theories for semantic memory and semantic composition. For semantic memory, the H&S approach models the ATL as a hidden 'hub' layer in a deep recurrent network, with reciprocal connections to 'spokes' layers each of which represents different object properties across sensorimotor and verbal modalities (Rogers et al., 2004; Lambon Ralph et al., 2010, 2017; Jackson et al., 2021). Input to any of these layers activates the corresponding pattern across the other layers (auto-associator) via the hidden hub layer, allowing for the extraction of transmodal and transtemporal statistical structures latent in verbal and non-verbal experiences across the lifespan. For semantic composition, computational models at the phrasal level (Landauer & Dumais, 1997; Clark & Pulman, 2007; Mitchell & Lapata, 2008, 2010) represent word meanings as vectors in high-dimensional spaces and implement composition as operations over constituent word vectors to derive phrase- and sentence-level vectors. Other models for sentence processing, such as the Sentence Gestalt models (McClelland et al., 1989; St. John & McClelland, 1990), conceive of sentence processing not as the building up of compositional structures. Instead, each word in a sentence serves as a cue to constrain and update a probabilistic representation about an event.

A recent model offers a framework that converges aspects of these models of semantic memory and semantic composition. Hoffman et al. (2018) proposed a neural network model that simultaneously (1) associates concrete words with sensorimotor features and (2) predicts upcoming words from context. Context sensitivity is captured through a recurrent architecture that buffers the hub unit's own state at a previous timestep. With a single set of "hub" units, the model can acquire semantic representations of concrete and abstract concepts as well as

their context dependence under the computational pressures of (1) and (2). This underlines the insight that conceptual semantic representations are not only captured by multimodal experiential properties but also governed by distributional statistics (Vigliocco et al., 2009).

The present study also paves the way for models of semantic cognition to better understand conceptual processing beyond simple adjective-noun phrases. Incorporating phrasal-level manipulations into models of semantic cognition could broaden their empirical coverage from single words through to time-extended coherent verbal narratives. A straightforward next step would be to extend the present approach to explore the conceptual processing of other linguistic elements, such as adjectives and verbs. For example, neuropsychological evidence from SD patients suggests that ATL similarly provides a representational basis for event concepts (denoted by verbs) and entity concepts (denoted by nouns) alike (Patterson et al., 2001; Bonner et al., 2009). MEG investigations on the composition of verb phrases also implicated the ATL (Kim & Pylkkänen, 2019). The approach taken in this study is thus well-positioned for further linking between semantic memory and composition. Our results have shown that adjective and noun semantics were represented with distinct formats (Figure 5B&D). Future work could explore models of semantic representation of different phrase types (e.g., noun vs. verb phrases) to explain the representational differences observed in our study. Moreover, model-internal representations and neural representations in humans could be linked for a potential mechanistic understanding of semantic cognition. Indeed, this is an approach that is commonplace is in vision neuroscience (e.g., Cichy et al., 2014; Kietzmann et al., 2019) and has recently been applied to studies of semantic cognition (Rogers et al., 2021).

Ultimately, we construct meaning out of rich, multimodal communicative situations and contexts. Such a unified framework for representing and integrating conceptual knowledge over time paves the way for new and exciting research questions about the nature of *natural* semantic cognition. For example, when communicating, we often use language together with non-linguistic cues, such as gestures (e.g., when giving directions, describing the size of an object). The conceptual processing of co-speech gestures has been extensively studied in brain and behavior (for review of general non-linguistic conceptual processing informed by the N400, see e.g. Kutas & Federmeier, 2011; for review of co-speech gesture processing, see e.g. Özyürek, 2014). Further, semantic composition in American Sign Language has been also shown to drive the left ATL (Blanco-Elorrieta et al., 2018). Thus, future work could explore the neurocomputational bases of the integration of conceptual information from linguistic and non-linguistic sources.

Our analyses revealed right ATL involvement, in addition to the left, during the conceptual processing of phrases, raising interesting questions about the role of the right ATL. The right ATL responded more strongly to phrases than words, regardless of phrase concreteness, around 420-450 ms post noun onset (Figure 3). While the effect size was smaller in the right relative to the left ATL, we suggest that the similar response profiles imply shared processes across the ATLs. Indeed, the proposal that both ATLs are functionally unified during conceptual processing has previously been advanced (Patterson et al., 2007; Schapiro et al., 2013; Lambon Ralph, 2014), with graded left-right asymmetries emerging from differential connectivity between the ATLs to different input/output modality systems (Lambon Ralph et al., 2001; Rice, Hoffman, et al., 2015; Rice, Lambon Ralph, et al., 2015; Schapiro et al., 2013). As the stimuli here were linguistic, the left ATL was thus preferentially engaged. The MEG-EEG results obtained here suggest that semantic memory and composition are unified both within *and across* the ATLs, complementing prior hemodynamic work and supporting an account of a bilateral-yet-graded, connectivity-constrained hub for semantic cognition.

Our analyses yielded new results on the timing of composition effects. The present composition effects localized between 390-660 ms post noun onset, which contrasts with prior

reports of effects around 200-250 ms post noun onset (Pylkkänen, 2019, 2020). Several study design differences may explain this discrepancy in timing. First, our stimuli were sampled across conceptual domains (e.g., objects, animals, emotions) while previous MEG studies tended to adopt concrete items (e.g., *red boat*) (Pylkkänen, 2019, 2020). The conceptual processing of items from different domains may place different demands on the ATL, leading to distinct effect timings. Second, our task was aimed at tapping into conceptual semantic processing that was appropriate for abstract concepts, which contrasts with previously used picture matching tasks (cf. Kochari et al., 2021). Task demand differences between a visual task that may encourage imagery and a comprehension task that promotes conceptual processing may explain effect timing differences.

Indeed, task demand differences due to study design changes may also have induced a change in the timings of our denotation effects. We found univariate effects around 380-440 ms and multivariate decoding effects around 370-450 ms post noun onset. This contrasts with Fritz and Baggio's (2020) EEG findings, which localized denotation effects around 500-800 ms. In our study, we used a diverse set of nouns (across conceptual domains) and adjectives (16 each of subsective and privative adjectives), while Fritz and Baggio (2020) used a smaller set of adjectives. Second, we adopted a comprehension task tapping into the processing of conceptual aspects of the nouns (e.g., for "genuine crown", the question was "Is it cheap?"). This contrasts with the task in Fritz and Baggio (2020), which directs the participant's attention to the modal aspects of the expression (e.g., for "fake president", the question was "Is it actually a president?"). These differences in task design may have contributed to the different effect timings without eliciting differential responses in semantic control regions (the left IFG and PTL; Jackson, 2021; Jackson et al., 2021; Lambon Ralph et al., 2017) across conditions. Therefore, one question that arises is how control systems contribute to conceptual processing during phrase comprehension.

Data and Code Availability

The data will be made available at http://www.mrc-cbu.cam.ac.uk/publications/open444data. Code and analysis pipelines will be made available at https://github.com/rmc-law/FakeDiamond.

Author Contributions

R.M.C.L.: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Writing – Original Draft, Visualization, Project Administration, Data Curation
O.H.: Conceptualization, Methodology, Resources, Writing – Review & Editing, Supervision, Project Administration, Funding Acquisition
M.A.L.R: Conceptualization, Methodology, Resources, Supervision, Writing – Review & Editing, Project Administration, Funding Acquisition
All authors reviewed and approved the final manuscript.

Funding

This work and the corresponding author (R.M.C.L.) were supported and funded by the Bill & Melinda Gates Foundation, Seattle, WA, and Gates Cambridge Trust (Grant Number: OPP1144). This study was supported by a programme grant to M.A.L.R. from the Medical Research Council (grant no. MR/R023883/1), and Medical Research Council intramural funding (no. MC_UU_00005/18).

Declaration of Competing Interest

The authors declare no competing interests.

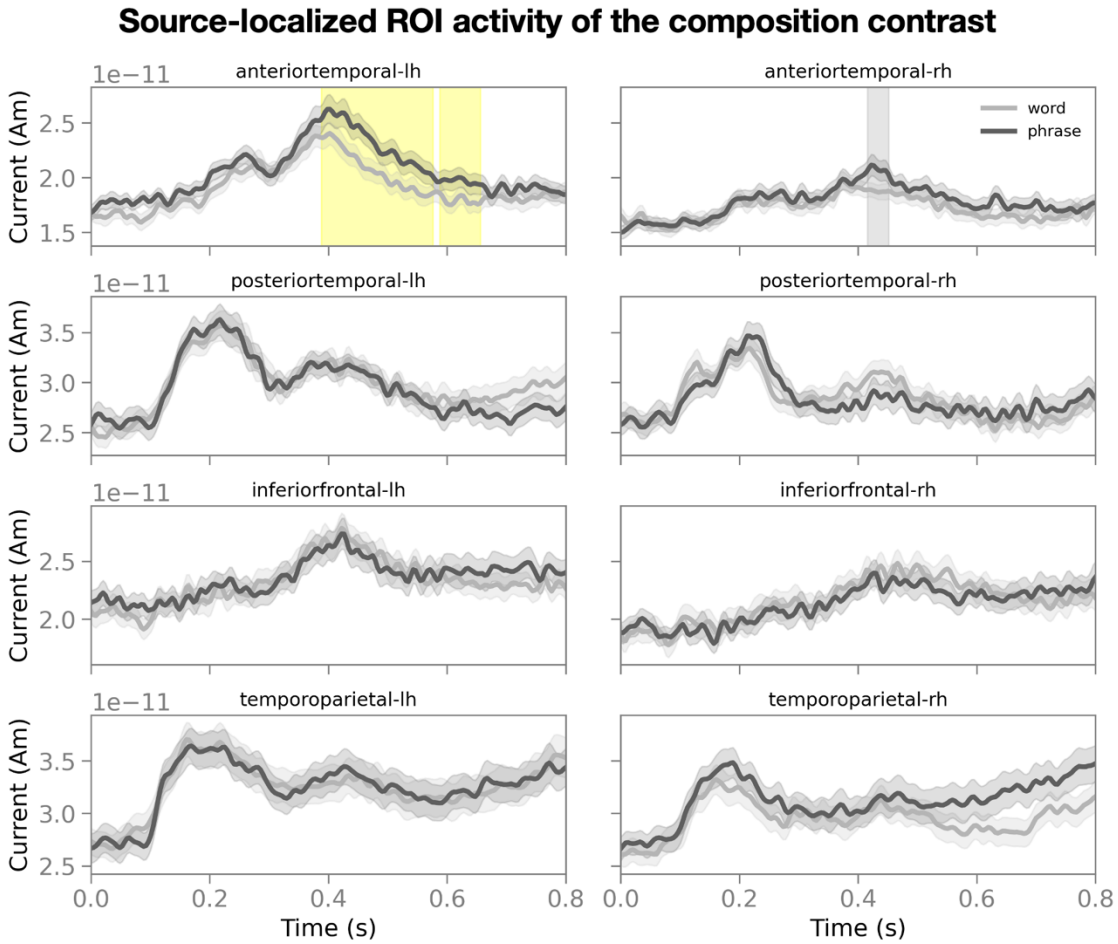
Acknowledgements

We are grateful to Clare Cook, Lucy MacGregor, Steve Eldridge, Máté Aller, Rebecca Williams, Fatih Serin, Marius Mada, and Karen Kabakulu for their assistance with data collection and also to Tobias Goehring and Lidea Shahidi for their advice on data analysis.

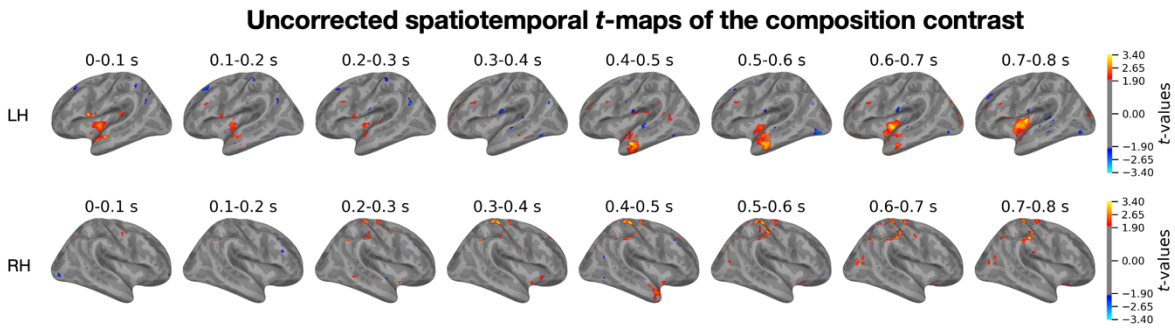
Author notes

Open access: For the purpose of open access, the UKRI-funded authors O.H. and M.A.L.R. have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission.

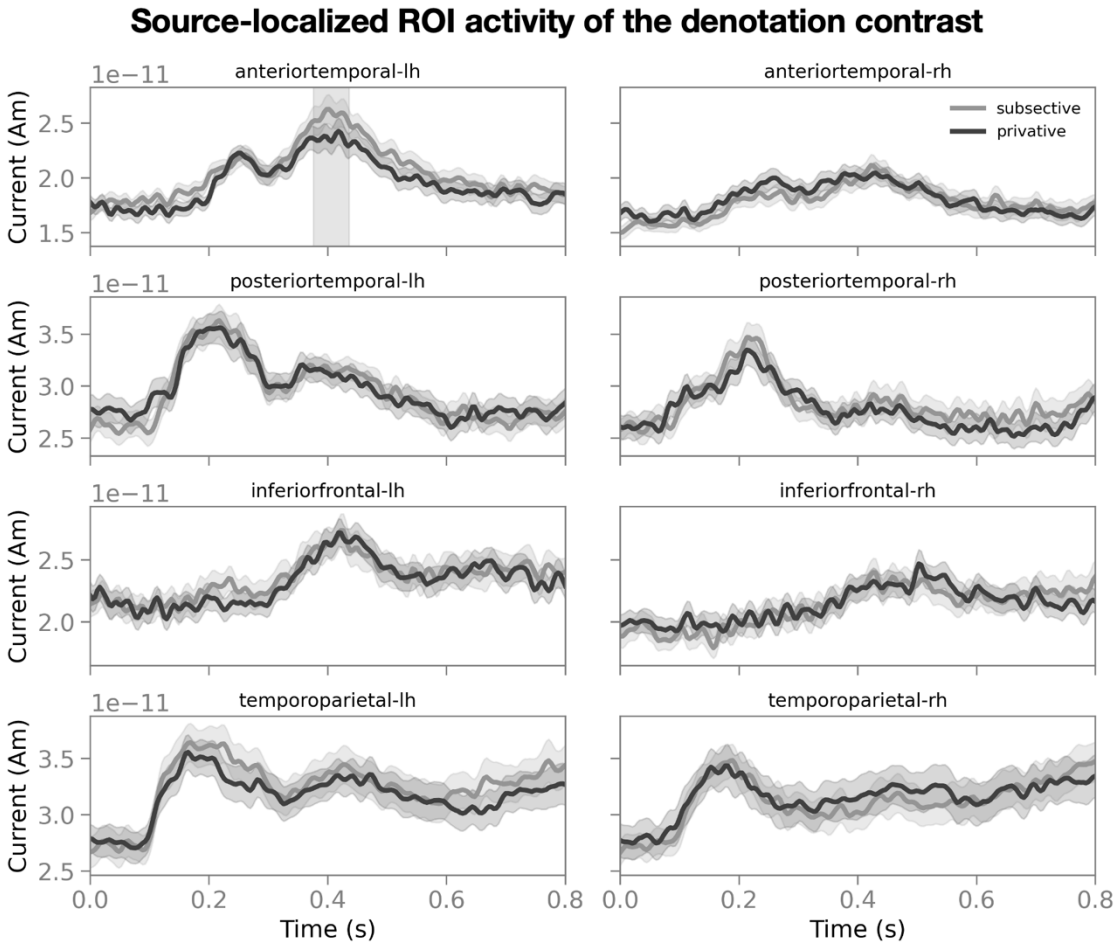
Supplementary Material



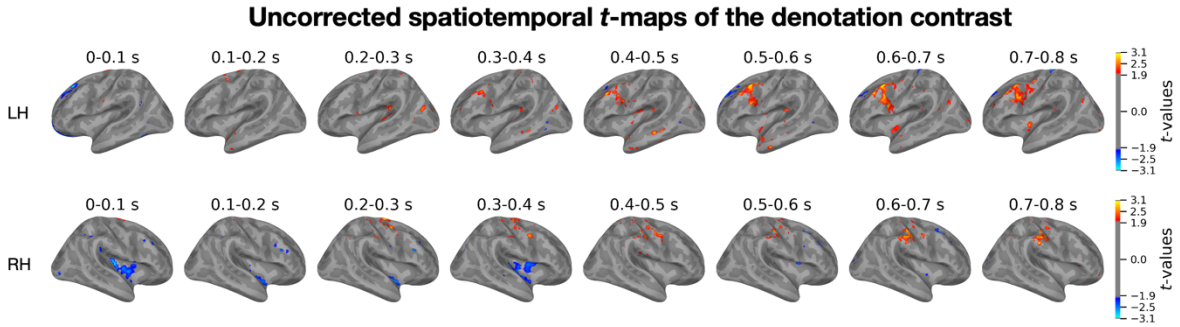
Supplementary Figure 1. Source-localized neural activity time series for each region of interest for subsective phrases vs. single words. Solid time series represent group-level averages, and shaded area represent one within-subjects SEM from mean. Yellow vertical spans demarcate cluster extents that correspond to statistically significant effects, while grey vertical spans correspond to marginal effects (see main text).



Supplementary Figure 2. Spatiotemporal (uncorrected) t-maps of composition effects (phrase – word) across both the left and right hemispheres. Each brain model depicts t-maps binned with each 100 ms wide time window.



Supplementary Figure 3. Source-localized neural activity time series for each region of interest for subsective phrases vs. privative phrases. Solid time series represent group-level averages, and shaded area represent one within-subjects SEM from mean. Grey vertical spans demarcate cluster extents that correspond to marginal effects (see main text).



Supplementary Figure 4. Spatiotemporal (uncorrected) t-maps of denotation effects (subsective – privative) across both the left and right hemispheres. Each brain model depicts t-maps binned with each 100 ms wide time window.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bemis, D. K., & Pylkkänen, L. (2011). Simple Composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *Journal of Neuroscience*, *31*(8), 2801–2814. https://doi.org/10.1523/JNEUROSCI.5003-10.2011

Blanco-Elorrieta, E., Kastner, I., Emmorey, K., & Pylkkänen, L. (2018). Shared neural correlates for building phrases in signed and spoken language. *Scientific Reports*, *8*(1), 5492. https://doi.org/10.1038/s41598-018-23915-0

Bonner, M. F., Vesely, L., Price, C., Anderson, C., Richmond, L., Farag, C., Avants, B., & Grossman, M. (2009). Reversal of the concreteness effect in semantic dementia. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643290903512305

Brodbeck, C., Das, P., Gillis, M., Kulasingham, J. P., Bhattasali, S., Gaston, P., Resnik, P., & Simon, J. Z. (2023). Eelbrain, a Python toolkit for time-continuous analysis with temporal response functions. *eLife*, *12*, e85012. https://doi.org/10.7554/eLife.85012

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Călinescu, L., Ramchand, G., & Baggio, G. (2023). How (not) to look for meaning composition in the brain: A reassessment of current experimental paradigms. *Frontiers in Language Sciences*, *2*. https://doi.org/10.3389/flang.2023.1096110

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

Clark, S., & Pulman, S. (2007). Combining symbolic and distributional models of meaning. *Proceedings of the AAAI Spring Symposium on Quantum Interaction*, 52–55.

Coulson, S., & Fauconnier, G. (1999). Fake guns and stone lions: Conceptual blending and privative adjectives. *Cognition and Function in Language*, 143–158.

Coutanche, M., Solomon, S., & Thompson-Schill, S. (2019). Conceptual combination. In *The Cognitive Neurosciences* (6th ed.).

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Del Pinal, G. (2015). Dual Content Semantics, privative adjectives, and dynamic compositionality. *Semantics and Pragmatics*, *8*, 7:1-53. https://doi.org/10.3765/sp.8.7

Engemann, D. A., & Gramfort, A. (2015). Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals. *NeuroImage*, *108*, 328–342. https://doi.org/10.1016/j.neuroimage.2014.12.040

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (Third). Sage. https://www.john-fox.ca/Companion/

Fritz, I., & Baggio, G. (2020). Meaning composition in minimal phrasal contexts: Distinct ERP effects of intensionality and denotation. *Language, Cognition and Neuroscience*, *35*(10), 1295–1313. https://doi.org/10.1080/23273798.2020.1749678

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

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Hansen, P. C., Kringelbach, M. L., & Salmelin, R. (Eds.). (2010). *MEG: An introduction to methods*. Oxford University Press.

Hoffman, P. (2016). The meaning of ‘life’ and other abstract words: Insights from neuropsychology. *Journal of Neuropsychology*, *10*(2), 317–343. https://doi.org/10.1111/jnp.12065

Hoffman, P., McClelland, J. L., & Lambon Ralph, M. A. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological Review*, *125*(3), 293. https://doi.org/10.1037/rev0000094

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*(4), 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5

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

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

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*, 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030

Kamp, J. A. W. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal Semantics of Natural Language* (pp. 123–155). Cambridge University Press; Cambridge Core. https://doi.org/10.1017/CBO9780511897696.011

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

Kim, S., & Pylkkänen, L. (2019). Composition of event concepts: Evidence for distinct roles for the left and right anterior temporal lobes. *Brain and Language*, *188*, 18–27. https://doi.org/10.1016/j.bandl.2018.11.003

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210.

Kochari, A. R., Lewis, A. G., Schoffelen, J.-M., & Schriefers, H. (2021). Semantic and syntactic composition of minimal adjective-noun phrases in Dutch: An MEG study. *Neuropsychologia*, *155*, 107754. https://doi.org/10.1016/j.neuropsychologia.2021.107754

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lalisse, M., & Asudeh, A. (2015). *Distinguishing intersective and non-intersective adjectives in compositional distributional semantics* [Master’s dissertation]. University of Oxford.

Lambon Ralph, M. A. (2014). Neurocognitive insights on conceptual knowledge and its breakdown. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120392. https://doi.org/10.1098/rstb.2012.0392

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

Lambon Ralph, M. A., McClelland, J. L., Patterson, K., Galton, C. J., & Hodges, J. R. (2001). No Right to Speak? The Relationship between Object Naming and Semantic Impairment:Neuropsychological Evidence and a Computational Model. *Journal of Cognitive Neuroscience*, *13*(3), 341–356. https://doi.org/10.1162/08989290151137395

Lambon Ralph, M. A., Sage, K., Jones, R. W., & Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences*, *107*(6), 2717–2722. https://doi.org/10.1073/pnas.0907307107

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Lenth, R. V. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://rvlenth.github.io/emmeans/

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. https://doi.org/10.3758/BF03210951

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

Matchin, W., Brodbeck, C., Hammerly, C., & Lau, E. (2019). The temporal dynamics of structure and content in sentence comprehension: Evidence from fMRI-constrained MEG. *Human Brain Mapping*, *40*(2), 663–678.

McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes*, *4*(3–4), SI287–SI335. https://doi.org/10.1080/01690968908406371

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., THE GOOGLE BOOKS TEAM, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182. https://doi.org/10.1126/science.1199644

Mitchell, J., & Lapata, M. (2008). Vector-based models of semantic composition. *Proceedings of ACL-08: HLT*, 236–244. https://aclanthology.org/P08-1028

Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*, *34*(8), 1388–1429. https://doi.org/10.1111/j.1551-6709.2010.01106.x

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

Nayak, N., Kowarsky, M., Angeli, G., & Manning, C. D. (2014). *A dictionary of nonsubsective adjectives* (CSTR 2014-04). Department of Computer Science, Stanford University.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130296. https://doi.org/10.1098/rstb.2013.0296

Paivio, Allan. (1971). *Imagery and verbal processes.* Holt, Rinehart and Winston.

Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, *108*(6), 2522–2527.

Partee, B. H. (2010). Privative adjectives: Subsective plus coercion. In R. Bauerle, U. Reyle, & T. Zimmermann (Eds.), *Presuppositions and discourse: Essays offered to Hans Kamp* (1st ed., Vol. 21, pp. 273–285). Emerald Group Publishing.

Patterson, K., Lambon Ralph, M. A., Hodges, J. R., & McClelland, J. L. (2001). Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia*, *39*(7), 709–724. https://doi.org/10.1016/S0028-3932(01)00008-2

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*(2), 184–187. https://doi.org/10.1016/0013-4694(89)90180-6

Pylkkänen, L. (2019). The neural basis of combinatory syntax and semantics. *Science*, *366*(6461), 62. https://doi.org/10.1126/science.aax0050

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*(1791), 20190299. https://doi.org/10.1098/rstb.2019.0299

Rabagliati, H., Doumas, L. A. A., & Bemis, D. K. (2017). Representing composed meanings through temporal binding. *Cognition*, *162*, 61–72. https://doi.org/10.1016/j.cognition.2017.01.013

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*(9), Article 9. https://doi.org/10.1038/s41562-018-0406-4

Rice, G. E., Hoffman, P., & Lambon Ralph, M. A. (2015). Graded specialization within and between the anterior temporal lobes. *Annals of the New York Academy of Sciences*, *1359*(1), 84–97. https://doi.org/10.1111/nyas.12951

Rice, G. E., Lambon Ralph, M. A., & Hoffman, P. (2015). The Roles of Left Versus Right Anterior Temporal Lobes in Conceptual Knowledge: An ALE Meta-analysis of 97 Functional Neuroimaging Studies. *Cerebral Cortex*, *25*(11), 4374–4391. https://doi.org/10.1093/cercor/bhv024

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

Rogers, T. T., Hocking, J., Noppeney, U., Mechelli, A., Gorno-Tempini, M. L., Patterson, K., & Price, C. J. (2006). Anterior temporal cortex and semantic memory: Reconciling findings from neuropsychology and functional imaging. *Cognitive, Affective, & Behavioral Neuroscience*, *6*(3), 201–213. https://doi.org/10.3758/CABN.6.3.201

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, *111*(1), 205.

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), e13335. https://doi.org/10.1111/psyp.13335

Schapiro, A. C., McClelland, J. L., Welbourne, S. R., Rogers, T. T., & Lambon Ralph, M. A. (2013). Why Bilateral Damage Is Worse than Unilateral Damage to the Brain. *Journal of Cognitive Neuroscience*, *25*(12), 2107–2123. https://doi.org/10.1162/jocn_a_00441

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 82–102. https://doi.org/10.1037/0278-7393.9.1.82

St. John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, *46*(1), 217–257. https://doi.org/10.1016/0004-3702(90)90008-N

Taulu, S., & Kajola, M. (2005). Presentation of electromagnetic multichannel data: The signal space separation method. *Journal of Applied Physics*, *97*(12), 124905. https://doi.org/10.1063/1.1935742

Taylor, J. E., Beith, A., & Sereno, S. C. (2020). LexOPS: An R package and user interface for the controlled generation of word stimuli. *Behavior Research Methods*, *52*(6), 2372–2382. https://doi.org/10.3758/s13428-020-01389-1

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247. https://doi.org/10.1515/LANGCOG.2009.011