

Tracking competition and cognitive control during language comprehension with multi-voxel pattern analysis



Elizabeth Musz*, Sharon L. Thompson-Schill

Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104, United States

ARTICLE INFO

Article history:

Received 11 January 2016

Revised 26 October 2016

Accepted 14 November 2016

Keywords:

Homonyms

Pattern similarity

Left ventrolateral prefrontal cortex

Semantic memory

Lexical ambiguity

MVPA

ABSTRACT

To successfully comprehend a sentence that contains a homonym, readers must select the ambiguous word's context-appropriate meaning. The outcome of this process is influenced both by top-down contextual support and bottom-up, word-specific characteristics. We examined how these factors jointly affect the neural signatures of lexical ambiguity resolution. We measured the similarity between multi-voxel patterns evoked by the same homonym in two distinct linguistic contexts: once after subjects read sentences that biased interpretation toward each homonym's most frequent, dominant meaning, and again after interpretation was biased toward a weaker, subordinate meaning. We predicted that, following a subordinate-biasing context, the dominant yet inappropriate meaning would nevertheless compete for activation, manifesting in increased similarity between the neural patterns evoked by the two word meanings. In left anterior temporal lobe (ATL), degree of within-word pattern similarity was positively predicted by the association strength of each homonym's dominant meaning. Further, within-word pattern similarity in left ATL was negatively predicted by item-specific responses in a region of left ventrolateral prefrontal cortex (VLPFC) sensitive to semantic conflict. These findings have implications for psycholinguistic models of lexical ambiguity resolution, and for the role of left VLPFC function during this process. Moreover, these findings demonstrate the utility of item-level, similarity-based analyses of fMRI data for our understanding of competition between co-activated word meanings during language comprehension.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

The field of psycholinguistics explains the resolution of lexical ambiguity as the consequence of selection between co-activated and competing interpretations of an ambiguous word. This view is akin to how researchers in the fields of perception, attention, and memory conceive of selection; namely, that it is a consequence of both bottom-up and top-down signals that drive competitive interactions between incompatible representations. In the present study, we take advantage of newly developed fMRI analysis techniques that have been usefully deployed to study the factors that influence selection and conflict resolution in domains of attention (e.g., Kamitani & Tong, 2005; Reddy, Kanwisher, & VanRullen, 2009) and memory (e.g., Kuhl, Rissman, Chun, & Wagner, 2011), and apply them for the first time to track competitive interactions during language comprehension. For instance, when readers must select a weaker, subordinate meaning of an ambiguous word (e.g.,

a river “bank”) over a stronger, dominant interpretation (e.g., a money “bank”), how (and where) does the resolution of this competition manifest in neural signals?

One useful approach for identifying interference from a task-irrelevant, competing response is to look for lingering “traces” of it in spatially distributed neural response patterns using multi-voxel pattern analyses (MVPA) of fMRI data. To accomplish this, researchers first measure the multi-voxel pattern (MVP) of activity evoked by a stimulus item, and then render this item irrelevant through a task manipulation. They then measure the MVPs elicited by another stimulus item that is somehow associated with the now-irrelevant stimulus, and determine the extent to which the MVPs evoked during the updated item resemble the responses that were evoked during the now-irrelevant, original item. In the episodic memory domain, researchers have used this technique to quantify competition during targeted memory retrieval, where the same cue simultaneously elicits two associated memories, although one of the associates is task-irrelevant (e.g., Kuhl, Bainbridge, & Chun, 2012; Wimber, Alink, Charest, Kriegeskorte, & Anderson, 2015). Similarly, in a study of event comprehension, Hindy, Solomon,

* Corresponding author.

E-mail address: muszeliz@sas.upenn.edu (E. Musz).

Altmann, and Thompson-Schill (2015) examined whether MVPs reflected the co-activation of two mutually exclusive states of the same object.

These studies have revealed that the degree of interference from the inappropriate representations, as manifested by their presence in MVPs in posterior cortical regions, was inversely predicted by increased recruitment of prefrontal cortex (PFC). We propose that PFC serves a domain-general role in biasing selection of task-relevant representations over competing alternatives. In the present study, we extend this proposal to the domain of lexical ambiguity resolution, and predict that PFC will similarly support the selection of MVPs evoked by subordinate, context-appropriate homonym meanings over dominant, context-inappropriate meanings.

1.1. Role of left ventrolateral prefrontal cortex in lexical ambiguity resolution

When comprehending everyday text and speech, the vast majority of words that we encounter have some degree of fluidity in their meaning, such that a single word might refer to one of several different meanings each time it is invoked. The sentence context in which a word is embedded serves as a critical cue to the word's intended meaning. Although context serves an irrefutable role in resolving this ambiguity, the relative scope and timing of its influence is largely unresolved. How (and when) do contextual factors influence word comprehension? In order to gain traction on these questions, numerous psycholinguistic experiments have investigated the online comprehension of lexically ambiguous words, such as homographic homophones. For these words (hereafter called homonyms), the same phonemic and orthographic markers refer to two or more distinct and unrelated meanings.

Because several meanings are associated with a single word form, even context-inappropriate, alternative meanings can be inadvertently activated upon encountering a homonym. Readers and listeners must rapidly select the appropriate referent at the expense of all other possible meanings, which may require resolving competition between co-activated referents. One candidate brain region for enabling a top-down bias toward context-appropriate representations is the left VLPFC (ventrolateral prefrontal cortex). In previous fMRI investigations, left VLPFC is consistently recruited during the presentation of sentences that contain homonym words, relative to unambiguous single-sense words (e.g., Hoening & Scheef, 2009; Rodd, 2005; Rodd et al., 2012; Vitello, Warren, Devlin, & Rodd, 2014). In addition, VLPFC activity (in particular, the left-lateralized inferior frontal gyrus and inferior frontal sulcus) increases when sentences bias interpretation toward (i.e., invoke) a homonym's subordinate meaning, relative to its dominant meaning (Zempleni, Renken, Hoeks, Hoogduin, & Stowe, 2007). Left VLPFC response is greatest for subordinate-biased "polarized" homonyms, whose subordinate meanings exhibit the weakest associations to the word form (Mason & Just, 2007). This response profile is consistent with the role of a modulatory mechanism that biases the interpretation of ambiguous words, either by boosting selection of the context-appropriate meaning, dampening selection of the inappropriate meaning, or some combination of the two.

1.2. Role of left ventrolateral prefrontal cortex in domain-general conflict resolution

More generally, beyond the domain of lexical ambiguity, this same region is consistently recruited during the resolution of competition amongst conflicting, co-activated representations (e.g.,

Hindy, Altmann, Kalenik, & Thompson-Schill, 2012; January, Trueswell, & Thompson-Schill, 2009; Thompson-Schill, Bedny, & Goldberg, 2005). In fact, the act of selecting a weaker word meaning amidst interference from a competing, stronger meaning has much in common with the processes involved in the Stroop task (MacLeod, 1991; Stroop, 1935). During incongruent trials of Stroop color-word interference task, subjects must respond according to one stimulus dimension (i.e., the word's display color) and ignore a stronger yet task-irrelevant dimension of that same stimulus that would yield an incorrect response (i.e., the color referred to by the stimulus word). Whether selecting a weak, subordinate meaning of a homonym word during lexical ambiguity resolution or reporting a stimulus words' display color instead of its name, in both cases, subjects must select between two simultaneous and mutually exclusive representations. To examine the functional and anatomical correspondences between lexical ambiguity resolution and domain-general cognitive control processes, we functionally localized subject-specific, conflict-sensitive regions of left VLPFC using a Stroop interference paradigm.

1.3. The current study

An extensive body of psycholinguistic research indicates that the competition between potential homonym meanings is greatest when the supporting context biases readers toward the selection of a subordinate referent that is only weakly associated with the word form (e.g., river-bank) (Duffy, Morris, & Rayner, 1988; Swaab, Brown, & Hagoort, 2003; Swinney, 1979). In order to resolve this conflict between co-activated alternatives, the reader must select the subordinate yet context-appropriate meaning over the dominant yet context-inappropriate meaning. What are the neural systems that support this process? Further, what neural and psychological factors influence the degree to which a dominant, inappropriate meaning is activated? To address these questions, we tracked the competition between homonym meanings as it unfolds in the brain.

We reason that dominant and subordinate meanings should evoke distinct neural responses in regions of the brain that are sensitive to variations in lexical-semantic information. To index competition between the two meanings, we computed the similarity between their corresponding neural patterns of activation. In particular, we measured the MVPs elicited while subjects first thought about a homonym's dominant meaning, and later on, its subordinate meaning. We then examined how the degree of competition between these neural responses (i.e., their neural similarity) varied across changes in meaning frequency; sentence context; and fluctuations in left VLPFC BOLD response.

We predicted that meaning frequency would positively predict the degree of competition. That is, the association strength between a homonym word form and its dominant meaning (i.e., its meaning frequency) should predict the similarity between the dominant-biased and subordinate-biased neural patterns, such that polarized homonyms should exhibit greater within-word neural similarity than more balanced homonyms, where the meaning frequencies of the dominant and subordinate meaning are relatively more equal. Secondly, we predicted that activity in left VLPFC would be associated with the top-down selection of the context-appropriate, subordinate meaning over the inappropriate, dominant meaning, and that this would manifest as decreased competition (i.e., less within-word neural similarity) during increases in left VLPFC response. As a secondary aim, we also investigated magnitude of BOLD response during sentence comprehension, and in particular, whether left VLPFC activity is modulated by the relative location of disambiguating sentence context.

2. Methods

2.1. Subjects

Thirteen right-handed, native English speakers (6 males), aged 20–29 years, participated in this study. Subjects were not taking any psychoactive medications and had no history of neurological disorders. All subjects had normal or corrected to normal vision. One additional subject was removed from analysis and replaced due to an unusually low response rate during the sentence-reading task (responded to 11% of trials, 4.4 standard deviations below the mean of all other subjects). Subjects were recruited from the University of Pennsylvania community. All subjects were paid \$20/hr and gave informed consent as approved by the University of Pennsylvania Institutional Review Board.

2.2. Stimuli

2.2.1. Main homonyms and meaning frequency (M1) scores

The main testing materials consisted of 30 ambiguous words in which the two most common meanings both refer to nouns (i.e., “ball”). These noun-noun homonyms were selected from a previous norming study that had tabulated the frequency counts of various meanings of several ambiguous words (Twilley, Dixon, Taylor, & Clark, 1994). In these norms, frequency scores for the most dominant word meaning (hereafter, M1) were computed by instructing behavioral subjects to generate a semantic associate for each ambiguous word. For each homonym, the authors determined the proportion of responses related to each possible meaning. In the present study, 30 of these homonyms were chosen to allow for a range of M1 scores across items ($M = 0.75$, $SD = 0.14$, Fig. 1). M1 scores are weakly correlated with log word frequency, $r = 0.29$, $t(28) = 1.66$, $p = 0.10$ (Brybaert & New, 2009).

2.2.2. Filler words

In addition to the main homonym stimuli, we included a set of single-sense nouns and additional noun-noun homonyms. First, 30 single-sense nouns (“synonyms”) were selected to match the dominant meaning of each main homonym. These synonyms were originally included to localize brain areas that exhibit similar MVPs in response to the dominant homonym meanings and their intended single-sense synonyms; however, this analysis failed to identify any reliable group-level effects. We will return to this null finding in the Discussion section. Second, to reduce the likelihood that subjects could predict the to-be-invoked meaning of a given homonym prior to sentence reading, we selected another 16 noun-noun homonyms and 12 single-sense nouns. Additional details about these filler word conditions are provided below.

2.2.3. Sentence stimuli

Each of the 30 main homonyms appeared in two different sentence conditions: once in a dominant-biasing context, and once in a subordinate-biasing sentence context. There were two types of

subordinate-biasing sentences: prior context (hereafter, sub-PC) and delayed context (hereafter, sub-DC). In sub-PC sentences, the homonym appeared near the end of the sentence, after the earlier words provide support for the subordinate homonym meaning. In sub-DC sentences, the homonym appeared early on in the sentence, such that the disambiguating contextual information was delayed until the end of the sentence (see Table 1). For the dominant-biasing sentences (hereafter, dom-PC), the homonym always appeared near the end of the sentence, preceded by words that supported the dominant meaning. Additionally, each dom-PC sentence was transformed into a single-sense sentence (hereafter single-syn) by replacing the homonym with its corresponding single-sense, synonymous noun. These four sentence conditions did not differ in letter length ($M = 37.2$, $SD = 3.6$), $F(3, 116) = 1.21$, $p > 0.3$ or number of words ($M = 6.9$, $SD = 0.89$), $F(3, 116) = 1.31$, $p > 0.2$. While all subjects received the same dom-PC and single-syn sentences, assignment of main homonym to either a sub-PC or sub-DC sentence was counterbalanced across subjects.

To ensure that all sentences could be read and adequately comprehended within the 3000 ms presentation duration employed during fMRI scanning, we first conducted a pilot study in which a separate group of behavioral subjects ($n = 6$) performed a self-paced reading task with these sentence stimuli. The sentence conditions were randomly interleaved, and each sentence was presented in isolation in the center of the display screen. Subjects were instructed to press a key once they were finished reading the sentence. To confirm that subjects semantically engaged with the sentences, 40% of the sentences were followed by comprehension questions that required subjects to make “yes” or “no” responses based on content from the immediately preceding sentence. Across stimulus conditions, subjects completed reading the sentences in less than 3000 ms ($M = 1871$ ms, $SD = 105$), and responded to the comprehension questions with well above chance performance ($M = 94.1$, $SD = 0.10$). To ensure that each individual sentence would be appropriate for the 3000 ms presentation time-frame, we applied conservative exclusion criteria: a sentence was removed or replaced if (1) it elicited a group average response time (RT) > 2500 ms or (2) the RT of any one subject exceeded 2800 ms.

2.3. Design overview

The primary goal of this procedure was to create conflict between two potential representations that might be retrieved upon the presentation of a homonym word. Findings from eye-movement studies, in which participants read sentences that contain an ambiguous word, indicate that readers require additional time to read disambiguating information that biases interpretation toward a homonym’s subordinate meaning (Rayner, 1998). We created a scenario to maximize the likelihood that subjects would retrieve the dominant, previously selected meaning of a homonym during the subsequent presentation of a subordinate-biasing context.

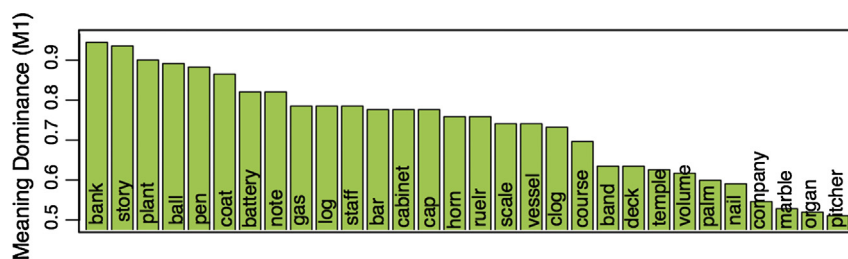


Fig. 1. The M1 (meaning dominance) score for the dominant meaning of the 30 main homonyms (Twilley et al., 1994).

Table 1

Example sentence conditions. Each sentence's respective homonym or single-sense synonym word is highlighted in bold above, but appeared in normal font during the experimental procedure. Dom-PC = dominant meaning, prior context; Sub-PC = subordinate meaning, prior context; Sub-DC = subordinate meaning, delayed context; Sing-Syn = single-sense word, synonym to dominant meaning; dom-DC = dominant meaning, delayed context.

Condition	Example sentence presentation
dom-PC	The fortune teller gazed into the crystal ball
sub-PC	The queen danced at her birthday ball
sub-DC	The ball was on the queen's birthday.
single-syn	The fortune teller gazed into the crystal orb
dom-DC	The trunk was filled with groceries

In the first half the experiment (runs 1–4), subjects read sentences that biased the interpretation of a main homonym toward its dominant meaning. After reading the sentence, the homonym was presented in isolation, and subjects were instructed to retrieve the word meaning which had been invoked in the immediately preceding sentence (i.e., the dominant meaning). In the second half of the experiment (runs 5–6), each main homonym then reappeared in a sentence that biased interpretation toward its subordinate meaning (either sub-PC or sub-DC, see Fig. 2). Subjects then again read each homonym word in isolation, this time retrieving the weaker meaning. Here, the question was whether the retrieval of the subordinate meaning would receive interference from the dominant, previously invoked meaning.

2.4. Trial sequences

We collected fMRI data during six acquisition runs comprising 134 trials. Each trial consisted of a 3000 ms sentence presentation, followed by 6000 ms fixation cross, and then the presentation of a single word from the preceding sentence (e.g., the main homonym) for 2500 ms (Fig. 2). Following the word presentation, a fixation cross was presented during a jittered ITI (500–12,500 ms). Within runs, trial orderings were randomized using Optseq2, an optimization program for sequencing trials in event-related experiments (<http://surfer.nmr.mgh.harvard.edu/optseq>).

Across runs 1–4, the 30 main homonyms each appeared in one dom-PC trial. In addition, a single-sense version of each dom-PC trial, in which the homonym was replaced with an unambiguous synonym (single-syn), also appeared in runs 1–4. The trial orders were pseudorandomized, such that a dom-PC trial never appeared in the same run as its single-syn counterpart. In runs 5–6, half of the main homonyms reappeared in a sub-PC trial, and the other

half appeared in a sub-DC trial. To balance the temporal distance between each homonym's dominant and subordinate presentations, subject trial sequences were yoked, such that the ordering for one subject was matched to another subject, but their sub-DC sentences were switched to sub-PC sentences, or vice-versa.

With these trial sequences, a homonym's invoked meaning could be predicted by the experiment half or a homonym's relative location in a sentence. To minimize these cues, we included sixteen filler homonyms that appeared in two different sentences, once in each experiment half. Both of its sentences biased interpretation toward the dominant meaning, and the homonym appeared early the sentence, such that the disambiguating context was delayed (i.e., dom-DC). In addition, six single-sense filler trials appeared in runs 5 and 6, such that half of the single-sense words appeared early on in their sentences, and the other half appeared later in the sentence. Runs 1–4 each consisted of 19 trials (5 min/run), and runs 5–6 each had 27 trials (7 min/run).

3. Procedure

3.1. Sentence-reading task

Stimuli were presented using E-Prime (Psychology Software Tools). Sentences appeared in the center of the screen in Arial font subtending approximately 0.5 degrees visual angle per letter. Subjects were instructed to respond via button press once they finished reading the sentence. After 3000 ms elapsed, the inter-stimulus interval (ISI) consisted of a centrally located fixation cross displayed for 6000 ms. Subjects responded to the majority of trials ($M = 85.1\%$, $SD = 16.7\%$), and each subject indicated during a post-scan debriefing session that they had adequate time to read each sentence. Across the main homonym sentence conditions (i.e., dom-PC, sub-DC, sub-PC) there were no significant differences in response times, $F(2,24) = 1.62$, $p = 0.22$. Mean response times (1817 ms) were consistent with the self-paced reading times from the pilot study (1871 ms), $t(10.13) = -0.27$, $p = 0.80$.

3.2. Semantic retrieval task

Following the sentence presentation and intervening fixation cross, a single word from the preceding sentence appeared on the screen for 2500 ms. Participants were instructed to think about the meaning of this word that was supported by the sentence context that they had just read. No behavioral measures were collected during this task.

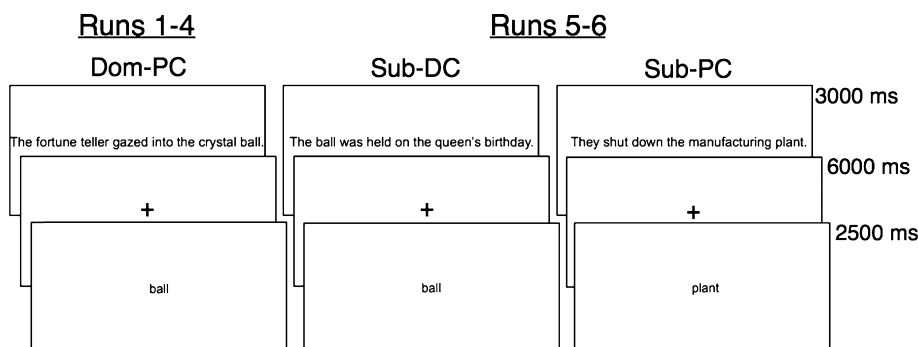


Fig. 2. Trial structure and condition sequences. Word stimuli first appeared in a sentence, followed by an isolated presentation of the targeted homonym or synonym. The main homonyms appeared in one Dom-PC sentence in runs 1–4, and in one sub-DC or one sub-PC sentence in runs 5–6. Subjects performed the sentence-reading task during the sentence presentations and the semantic retrieval task during the word presentations. Each semantic retrieval trial was followed by a jittered inter-trial interval for 500–12,500 ms during which a fixation cross was displayed.

3.3. Stroop interference task

After completing runs 1–6, subjects completed a single run of a Stroop color identification task (cf. Hindy et al., 2012, 2015). On each trial, subjects were presented with a single word and were instructed to press one of three response buttons that corresponded to the typeface color (i.e., blue, yellow, or green). The single word referred to either a color name (e.g., yellow, red) or a non-color, neutral noun (e.g., stage, tax, and farmer). Each word appeared for 1800 ms followed by a 1200 ms ITI. The conflict condition consisted of trials where the color name did not match the color of the typeface. In the neutral condition, the color name and typeface color matched, or a non-color, neutral noun was presented. Subjects responded correctly to 98.4% of Stroop trials. Response latencies for conflict trials ($M = 721$ ms, $SD = 186$ ms) were slower than responses to neutral trials ($M = 671$ ms, $SD = 191$ ms), $t(12) = 7.90$, $p < 0.001$). In a group-level, univariate contrast of conflict versus neutral trials, left VLPFC was reliably more responsive to Stroop conflict than adjacent brain regions. The anatomical location of the top 100 conflict-responsive voxels in left VLPFC was heterogeneous across subjects (Fig. 3).

3.4. fMRI data acquisition

Anatomical and functional data were collected on a 3T Siemens Trio system and a 32 channel array head coil. Anatomical data consisted of 160 slices of axial T1-weighted images with 1 mm isotropic voxels ($TR = 1620$ ms, $TE = 3.87$ ms, $TI = 950$ ms). Functional data included echo-planar fMRI collected in 44 axial slices and 3 mm isotropic voxels ($TR = 3000$ ms, $TE = 30$ ms). To approach steady state magnetization, twelve seconds preceded data acquisition in each functional run.

3.5. fMRI preprocessing

Data preprocessing and statistical analyses were performed with AFNI (Cox, 1996) and MATLAB scripts implemented in the Princeton MVPA Toolbox (Detre et al., 2006). Functional data were since interpolated for slice timing correction, aligned to the mean of all function images using a seventh-order polynomial interpolation, and co-registered to the structural data. Data were then smoothed with a 4 mm FWHM Gaussian kernel and z-normalized within each run.

3.6. Whole-brain regression analyses

We preformed two whole-brain analyses: a condition-level, univariate analysis, and an item-level, multi-voxel pattern (MVP) analysis. In both cases, a modified general linear model (Worsley & Friston, 1995) was fit to each subject's preprocessed data. Each trial segment was modeled with a canonical hemodynamic response function convolved with a boxcar that matched the duration of the trial segment (i.e., 3000 ms for each sentence, 6000 ms for each fixation ISI, and 2500 ms for each word). For the condition-level, univariate analysis, a binary regressor was included for each sentence and word condition (i.e., dom-DC; sub-DC; sub-PC; single-syn; and dom-DC). For the item-level MVP analysis, a unique regressor was included for each individual sentence and word presentation. For both models, scanning run and six motion parameters were modeled as covariates of no interest. For group-level, random-effects analyses, subject-level statistical maps were normalized to Talarach space. In order to correct for multiple comparisons, minimum cluster extent was determined using AFNI's 3dClustSim (version built December 9, 2015). For this correction, we first estimated the smoothness of the data using the residual time series data using AFNI's 3dFWHMx spatial autocorrelation function. Based on a voxel-level uncorrected alpha of 0.001 ($t = 4.29$), Monte Carlo simulations ($n = 50,000$) indicated a minimum cluster extent of 10 voxels for a cluster-level corrected alpha of 0.05.

3.7. ROI analysis: left VLPFC

Each Stroop-conflict ROI was anatomically constrained according to probabilistic anatomical atlases that were transformed into Talarach space (Eickhoff et al., 2005). Left VLPFC was defined as the combination of pars opercularis (BA 44), pars triangularis (BA 45), and the anterior half of the inferior frontal sulcus. Because the Stroop task entails multiple, distinct forms of conflict (e.g., motor response, task set, and color representation), this anatomical constraint allows for the selection of cortical areas that are most likely to be involved in the cognitive process of interest. The anatomical constraint to left VLPFC ensured that this ROI reflected conflict-related processing at the level of semantic representation (cf. Hindy et al., 2012). Across subjects, this left VLPFC anatomical ROI consisted of an average of 1024 voxels ($SD = 99$). Within these anatomical boundaries, the Stroop-conflict ROI was further limited according to each individual

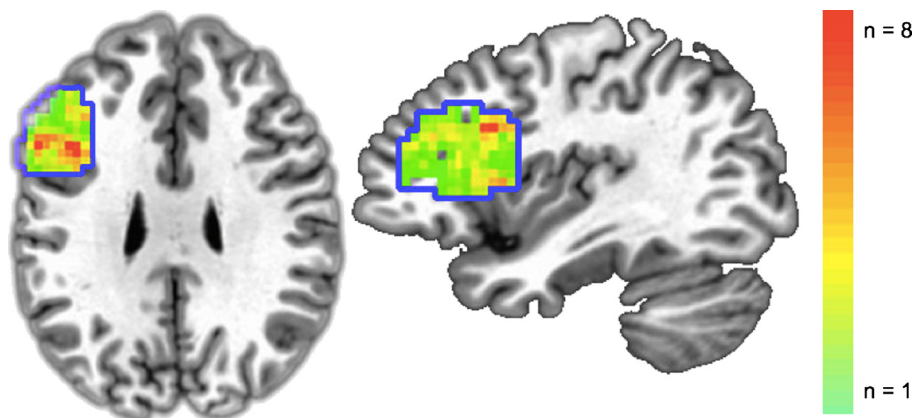


Fig. 3. Probabilistic overlap map of the subject-specific Stroop-conflict ROIs in left pVLPFC. Anatomical constraints of left VLPFC are outlined in blue. This anatomical ROI was transformed into each subject's native brain space. In each subject, we selected the 100 voxels which yielded the highest t -statistics in the contrast of conflict versus neutral trials during the Stroop task. For display purposes, these subject-level masks were transformed to standardized Talarach space and overlaid to create a group mask.

subject's functional data from the Stroop color-word interference task. Specifically, the ROI was confined to the 100 voxels that exhibited the highest t -statistics for the contrast of conflict versus neutral trials. This functional constraint ensures that the voxels included in this ROI were most sensitive to conflict on a subject-specific basis.

For the ROI-based regression analyses, voxel-wise activation values were averaged across the entire Stroop-conflict left VLPFC ROI in each subject. For the condition-level analysis, we tested the same contrasts described in the whole-brain analysis. For the item-level analysis, we measured the mean BOLD signal evoked during each main homonym's two word presentations (i.e., following its dominant- and subordinate-biasing sentences), and subtracted the average "dominant" response from the average "subordinate" response. This item-level measure serves as an index of the change in left VLPFC recruitment during the presentation of the dominant versus subordinate meaning of each main homonym.

3.8. Whole-brain multi-voxel pattern searchlight analysis

To assess the similarity of multi-voxel, item-specific responses evoked during each word presentation, we passed a spherical searchlight with a 3-voxel radius over each voxel in the brain (Kriegeskorte, Goebel, & Bandettini, 2006). (The main searchlight results were also confirmed when the searchlight size was increased to a 4-voxel radius). In each searchlight volume, MVP similarity was measured as the Pearson correlation between the multi-voxel responses evoked by the dominant versus subordinate word presentations of the same main homonym. In a subject-level, parametric analysis, we used M1 scores to predict the similarity between the MVPs evoked during each homonym's dominant and subordinate presentations. Here, we estimated a separate linear regression coefficient for each subject that predicted the MVP similarity of each homonym based on its M1 score. The resulting beta value was then assigned to each searchlight center. We then used 1-sample t -tests to determine the cross-subject reliability of the regression coefficients. This analysis is akin to entering Pearson correlation coefficients in a second-level analysis, instead of linear regression coefficients.

4. Results

4.1. Univariate results

4.1.1. Whole-brain analysis

In an exploratory, whole-brain analysis, we first contrasted the responses for the various sentence conditions. The contrast between sub-DC sentences versus dom-PC sentences yielded a large area of activation in left VLPFC, extending anterior and dorsal to the Stroop-conflict functional ROI. This cluster overlapped with a cluster resulting from the contrast of sub-DC sentences versus sub-PC sentences (Fig. 4). The coordinates and peak voxel values are listed in Table 2. The contrast of sub-PC versus dom-PC did not yield any reliable above-threshold activation, nor did the contrast of single-syn sentences versus any of the three homonym sentence conditions.

4.1.2. Stroop-conflict selective voxels in left VLPFC ROI

In addition to the whole-brain analysis, we compared the mean BOLD response for each sentence condition in each subject's top 100 Stroop-selective voxels in an anatomically constrained region of left VLPFC (Fig. 5). This analysis recapitulated the results that emerged at the whole brain level: mean left VLPFC response in the Stroop-conflict selective voxels was greater during the presentation of sub-DC sentences than sub-PC sentences, $t(12) = 4.20$, $p = 0.001$, and for sub-DC sentences versus dom-PC sentences $t(12) = 3.50$, $p = 0.004$. In addition, mean response was greater for sub-DC sentences versus single-syn sentences, $t(12) = 2.46$, $p = 0.03$.

4.2. Multi-voxel searchlight results

4.2.1. Role of meaning frequency

We used a whole-brain, multi-voxel searchlight analysis to examine the similarity between the MVPs evoked during the dominant-biased versus subordinate-biased version of the same homonym. In a group-level analysis, we performed a random-effects analysis using the statistical maps yielded by each subject's searchlight results, in which the linear regression coefficient for M1 was assigned to the searchlight centers. Across subjects, we identified a cluster of 21 searchlight volumes in left anterior

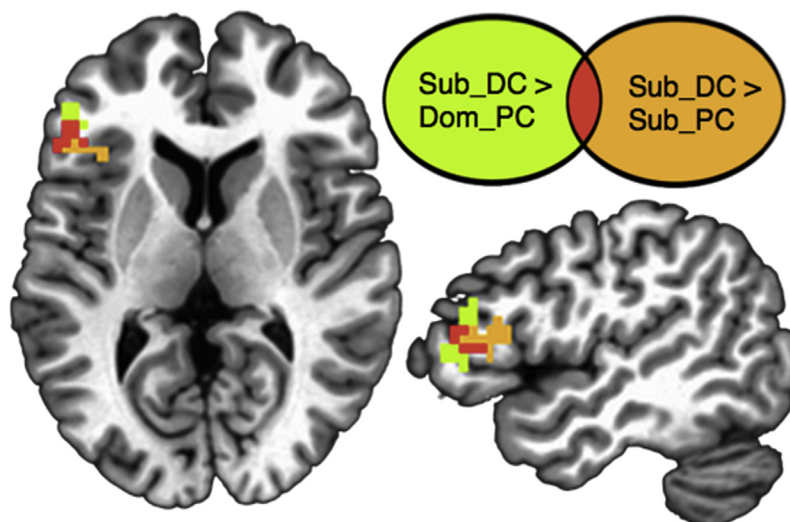


Fig. 4. Univariate whole-brain results for BOLD responses during the sentence-reading task. Subordinate-delayed context (Sub-DC) sentences elicited a greater response than both dominant-prior context (Dom-PC) and subordinate-prior context (Sub-PC) sentences in an overlapping area of left inferior frontal gyrus. Colored voxels depict areas with above-threshold activity in a cluster-corrected group-level analysis.

Table 2

Whole-brain, group level results. Responses in left inferior frontal gyrus increased during the presentation of subordinate-delayed context (sub-DC) sentences, relative to subordinate-prior context (sub-PC) and dominant-prior context (dom-PC) sentences. During the subsequent presentation of each sentence's homonym word, within-word multi-voxel pattern similarity positively correlated with meaning dominance in left anterior temporal lobe (left ATL). In addition, within-word pattern similarity in the peak left ATL searchlight sphere negatively predicted BOLD response in Stroop conflict-sensitive regions of left ventrolateral prefrontal cortex (IVLPFC).

Task	Effect	Location	Peak coordinates (x, y, z)	Cluster extent	Peak <i>t</i> -statistic
Sentence reading	BOLD response: sub-DC > sub-PC	Left inferior frontal gyrus	−46, 26, 8	17	$t(12) = 6.35, p < 0.001$
Sentence reading	BOLD response: sub-DC > dom-PC	Left inferior frontal gyrus	−46, 35, 11	10	$t(12) = 6.53, p < 0.0001$
Semantic retrieval	Within-word neural similarity: positively correlated with meaning dominance	Left anterior temporal lobe	−37, −7, −31	21	$t(12) = 5.45, p < 0.0001$
Semantic retrieval	Within-word neural similarity: negatively correlated with Stroop-conflict selective IVLPFC response	Left anterior temporal lobe			$t(12) = -3.14, p < 0.01$

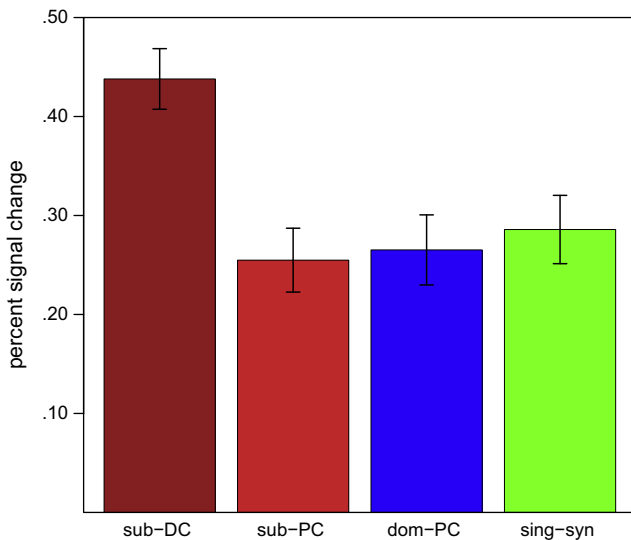


Fig. 5. Group-average responses during sentence comprehension in left posterior ventrolateral prefrontal cortex, limited to subject-specific, Stroop-conflict selective voxels. Comparisons between sentence conditions were performed within each subject. Sub-DC = subordinate meaning, delayed context; Sub-PC = subordinate meaning, prior context; Dom-PC = dominant meaning, prior context; Sing-Syn = single-sense word, synonym to dominant meaning.

temporal lobe (ATL) in which M1 scores reliably predicted the similarity between the MVPs evoked by the dominant- and subordinate-biased presentations of a main homonym (see Table 2 and Fig. 6a), $t(12) = 5.45, p = 0.0001$ (mean $r = 0.22, SD = 0.13$). In the MVPs sampled in these searchlight volumes, the greater the homonym's M1 score, the greater the similarity in the responses evoked by the two versions of the same ambiguous word. This relationship was positive in all 13 subjects (Fig. 6b). Follow-up analyses at the peak left ATL searchlight, in which homonyms were separated based on the relative location of the subordinate-biasing sentence context (i.e., sub-DC or sub-PC) indicate that this result holds when the analysis is limited to the main homonyms that had appeared in sub-DC sentences, $t(12) = 2.88, p = 0.01$, and marginally holds for the sub-PC homonyms alone as well $t(12) = 2.04, p = 0.06$.

Neural similarity was computed using Pearson's r , a similarity measure that is assumed to be largely independent of the absolute magnitude of univariate response. To confirm that the MVP similarity effects we observed in left ATL reveal information that is not redundant to univariate effects, we submitted the neural similarity values to a confirmatory, within-subject regression at the peak left ATL searchlight. For this regression analysis, we used four independent variables to predict M1 scores: neural similarity between the dominant and subordinate MVPs; mean univariate activity during the dominant retrieval; mean univariate activity during the subordinate retrieval; and the interaction between the mean univariate activity during each retrieval period (cf. Ritchey, Wing, LaBar, &

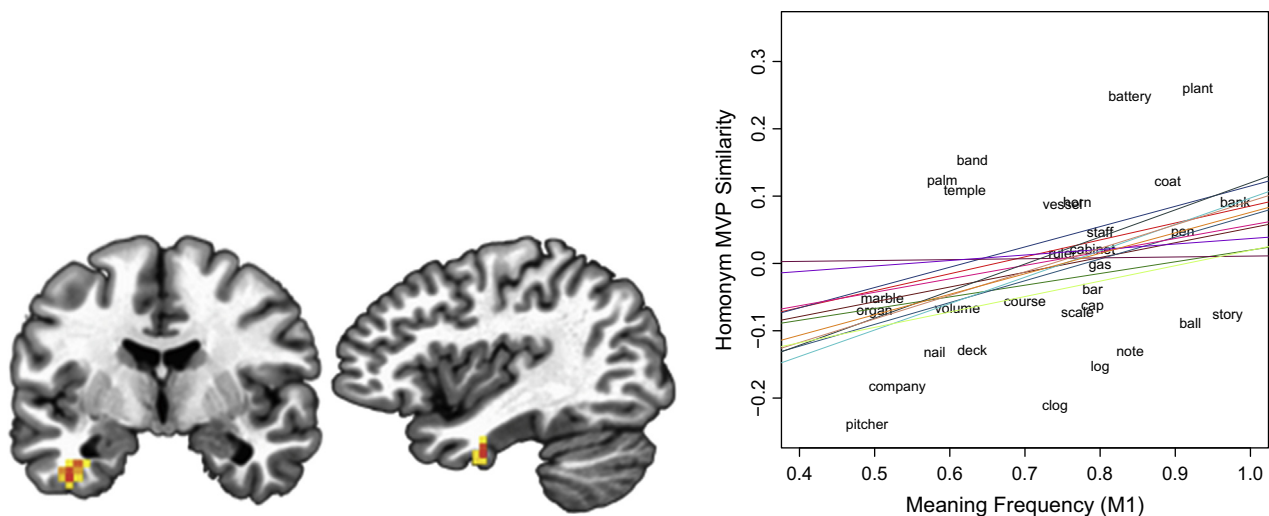


Fig. 6. (a) In left anterior temporal lobe, meaning frequency (M1) predicted the similarity between the neural patterns evoked during the semantic retrieval of a homonym's dominant and subordinate meanings. (b) The positive relationship between multi-voxel pattern (MVP) similarity and meaning frequency was present in all 13 subjects. The linear trend for each subject is depicted in a different color. Item-level results in a single subject are depicted in the background.

Cabeza, 2013; Wing, Ritchey, & Cabeza, 2015). Across subjects, the beta coefficient for MVP similarity continued to reliably predicted M1 scores, even with mean univariate response included in the model, $t(12) = 6.4, p < 0.001$ ($M = 0.21, SD = 0.12$). This confirmatory analysis minimizes the possibility that the searchlight results in left ATL are driven by mean activation differences.

In a follow-up analysis, we extracted the Pearson correlation coefficients for each main homonym at this peak left ATL searchlight center, and then used these values as a seed in a whole-brain analysis to predict changes in univariate response between the dominant versus subordinate word presentations. No reliable above-threshold activity emerged from this analysis.

4.2.2. Role of left VLPFC response

We also examined role the relationship between left VLPFC activity and dominant and subordinate MVP similarity. For this analysis, we selected the neural similarity values from the peak left ATL searchlight center where within-word neural similarity had exhibited the positive correlation with M1 scores. We then correlated changes in Stroop-selective left VLPFC response during the homonym presentations with the MVP similarities in this peak left ATL searchlight. Across subjects, increases in left VLPFC response from the dominant to subordinate word presentation reliably predicted decreases in the neural similarity between the subordinate and dominant word presentations of the same homonym in left ATL, $t(12) = -3.14, p = 0.01$ ($M = -0.07, SD = 0.08$). This relationship was negative in 10 out of 13 subjects.

To further investigate the effects of left VLPFC response on within-word neural similarity, we also performed an exploratory whole-brain searchlight analysis. Here, the change in left VLPFC response in subject-specific Stroop-conflict voxels between the subordinate versus dominant word presentation were used as predictors of MVP similarities in searchlights passed over the entire brain volume. This analysis failed to yield any reliable results at the whole-brain, group level.

4.2.3. Left anterior temporal lobe results: role of left VLPFC and meaning frequency

In a subject-level linear regression analysis, we predicted the neural similarity values observed in the peak left ATL searchlight by modeling separate covariates for M1 scores and change in left VLPFC response. Across subjects, the covariates for M1 and change in left VLPFC both reliably predicted neural similarity in left ATL, even when both covariates were simultaneously included in the model, $t(12) = 7.97, p = 0.0001$ for the M1 covariate, and $t(12) = -2.76, p = 0.02$ for the left VLPFC covariate. Moreover, M1 scores and changes in left VLPFC response were not reliably correlated across subjects, $t(12) = -0.23, p = 0.81$ ($M = -0.01, SD = 0.17$).

5. Discussion

Several neural and behavioral factors have been implicated in semantic ambiguity resolution, including left VLPFC response, homonym-level properties (i.e., meaning frequency), and sentence-level characteristics (i.e., the relative location of disambiguating context). We examined the role of these factors while tracking the outcome of ambiguity resolution using online, item-level neural measures. Our analyses revealed that these three factors each impact the neural correlates of lexical ambiguity resolution. In turn, we discuss each finding and the implications for psycholinguistic models of ambiguity resolution.

5.1. Univariate findings during sentence reading

We first examined changes in BOLD response while subjects read ambiguous noun-noun homonyms within sentence contexts.

A whole-brain analysis revealed that BOLD response in left VLPFC was modulated by meaning frequency, such that activity here was greater for subordinate-biasing versus dominant-biasing sentences. However, this effect was limited to subordinate-biasing sentences in which the disambiguating context was delayed (sub-DC). Additionally, in an overlapping set of voxels in left VLPFC, an effect of context position emerged for subordinate-biasing sentences, such that responses were greater when the disambiguating context followed the homonym (sub-DC) compared to when the context preceded it (sub-PC). This pattern of results was recapitulated in an fROI-based analysis, in which we selected subject-specific voxels in left VLPFC that were most responsive to conflict during a Stroop color-word interference task. This approach is important, because there have been suggestions that left VLPFC is a highly heterogeneous region, and subject-specific analyses are necessary to localize activity associated with the distinct process of interest (Fedorenko, Hsieh, Nieto-Castanon, Whitfield-Gabrieli, & Kanwisher, 2010).

Taken together, these findings confirm the role of left VLPFC in sentence reinterpretation and resolving competition between co-activated representations. The increased recruitment that we observed here is consistent with a scenario in which a frequency-based probabilistic choice is made between the alternative meanings, and then the meaning is updated if the selected interpretation does not fit with the subsequent disambiguating context (Zemleni et al., 2007).

5.2. Multivariate findings during semantic retrieval

In addition to examining neural activity during sentence reading, we also measured the neural activity that followed this disambiguation process, once the context had biased interpretation toward a particular homonym meaning. In previous work (Musz & Thompson-Schill, 2015), we have demonstrated the utility of within-item, cross-context neural similarity analyses by showing that the MVP similarity elicited by the same word across different presentations can be predicted by item-level semantic properties. In the present experiment, we employed sentence contexts to bias semantic retrieval toward one of two specific and distinct homonym meanings. We predicted that the neural representation evoked by the same word in the two different contexts would vary, such that these two different meanings would evoke variable neural patterns. Further, we examined the effects of switching the context (and hence the meaning) while holding the word form constant, such that a previously invoked meaning is rendered inappropriate and potentially distracting. Thus, retrieval of the subordinate meaning would require the subject to disregard a salient yet contextually inappropriate word meaning in favor of the weaker representation of the same word.

5.2.1. Meaning frequency predicts within-word neural similarity in left anterior temporal lobe

We first tested whether meaning frequency correlated with the extent to which subordinate-biased activity patterns resemble dominant-biased MVPs during retrieval of a subordinate meaning. A whole-brain searchlight analysis revealed that, in left ATL, the association strength of the dominant meaning (i.e., M1) predicted the degree of neural similarity between the dominant and subordinate-biased MVPs. Crucially, this effect emerged during the time period that followed the homonym's appearance in a sentence that biased interpretation toward its subordinate meaning. That is, even *after* the subordinate meaning had been supported via linguistic context, the neural patterns in left ATL still resembled those evoked by the dominant meaning. This finding adds to a growing literature on the role of meaning dominance during lexical ambiguity resolution. These investigations have largely found that

the dominant meaning of a homonym interferes with the selection of a subordinate homonym, and this competition between co-activated meanings leads to processing costs (Pacht & Rayner, 1993; Rayner, Pacht, & Duffy, 1994) and increased recruitment of left VLPFC (which we also observed during the presentation of sub-DC sentences). However, the majority of previous studies focused on the time interval during which a subject first encounters the subordinate-biased homonym (cf. Gorfein, 2001). In the present analysis, the neural pattern evoked by subordinate meaning was measured six seconds after this meaning had already been invoked in the preceding sentence. Thus, in addition to the competition that arises when a homonym word meaning is first accessed or reinterpreted, we found evidence of competition even after the word meaning has been resolved.

This finding indicates that even when the dominant meaning is rendered irrelevant by an earlier, subordinate-biasing sentence context, it nevertheless competes for activation. A host of previous behavioral research corroborates this finding. Several studies on reading times have revealed that subjects experience processing delays (manifested in increased reading times and regressive eye movements) while selecting in the subordinate meaning of a homonym, even when the supporting linguistic context has supported its interpretation (cf. Duffy et al., 1998; Pacht & Rayner, 1993) Sereno, O'Donnell, and Rayner (2006). This performance decrement, termed the “Subordinate Bias Effect” (SBE) has been demonstrated under several experimental conditions in which a previous context is provided to bias interpretation toward the subordinate meaning (e.g., paragraph titles, immediately preceding uses of the subordinate meaning, etc.).

Behavioral studies have found, however, that the SBE can in fact be eliminated by a strong subordinate-biasing preceding context, but only for ambiguous words that are only moderately biased (8–30% strength of the subordinate meaning). For polarized homonyms, in which the strength of the subordinate meaning was very weak (8% or less), the interference from the dominant meaning could not be fully eliminated (Wiley & Rayner, 2000). In a related study, Rodd, Johnsrude, and Davis (2012) investigated the extent to which lexical-semantic re-turning can rapidly occur. Subjects performed a free association task, in which they were presented with a homonym word and were instructed to generate a semantic associate. Twenty minutes beforehand, subjects in the primed condition listened to sentences that invoked the homonyms' subordinate meanings. Relative to unprimed subjects, the primed group was more likely to subsequently generate words related to the subordinate meanings. However, the priming effect was relatively modest: although the proportion of subordinate associates of polarized homonyms increased fivefold (e.g., from 2% to 10%), subjects were still far more likely to produce an associate of the dominant meaning. Taken together, these results suggest that even strong subordinate-biasing contexts cannot override the unintended dominant meaning if it has a very high frequency.

An eyetracking study by Huettig and Altmann (2007) provides a particularly striking demonstration of the interference from context-inappropriate, dominant homonym meanings. In a visual word paradigm, subjects viewed an array of four objects, where some of these objects depicted a homonym's subordinate meaning (e.g., a pig pen) and either its dominant meaning (a writing pen) or an object related in shape to the dominant meaning (e.g., a sewing needle). During the auditory presentation of a subordinate-biasing sentence context, fixations increased for the dominant competitor, and even for an object related in shape to the dominant referent, relative to unrelated control objects. Looks to these competitor objects can be interpreted as evidence that the dominant meaning was activated, despite the contextual support for the subordinate meaning.

Whereas those authors found evidence of transient, online activation of dominant meanings via eye fixations, we tracked the activation of homonym meanings as manifested in the similarity of their evoked neural signals. The neural similarity effects emerged in a left-lateralized subregion of the anterior temporal lobe. This area has been previously associated with increased recruitment during the retrieval of multiple ambiguous word meanings. In a recent study on homonym comprehension, Whitney, Jefferies, and Kircher (2011) found that BOLD activity in this same region was sensitive to the number of homonym meanings that were retrieved. Additionally, Snijders et al. (2009) reported increased activity in an overlapping region of left mid-inferior temporal gyrus (BA 20) while subjects read homonyms that were embedded in equibiasing sentence contexts, such that two alternative interpretations of the ambiguous word were equally plausible. In conjunction with our effect, these findings suggest that responses in this subregion of left ATL track the activation of several co-activated interpretations of ambiguous words.

These findings are also consistent with a host of previous research that points to a critical role for left ATL in semantic memory. This area's role in semantic processing has been established by convergent findings from patient studies, neuroimaging studies, and brain stimulation research (e.g., Patterson, Nestor, & Rogers, 2007; Visser, Jefferies, & Ralph, 2010; Visser & Ralph, 2011; Pobric, Jefferies, & Lambon Ralph, 2010). In fact, our identified searchlight cluster directly overlaps with a site recently identified as critical for semantic processing: Binney, Embleton, Jefferies, Parker, and Lambon Ralph (2010) found that BOLD response in this same subregion of left ATL increases while healthy subjects perform a synonym judgment task, and that Semantic Dementia patients with damage to this region exhibit impaired performance on the same task.

In light of the extant findings that implicate left ATL in conceptual processing, we suggest that the MVP similarities that we have identified here reflect the co-activation of the meanings associated with two alternative interpretations of the same homonymous word. However, we cannot conclusively attribute our effects to the activation of semantic information. In a preliminary, whole-brain analysis, we attempted to localize brain areas in which neural similarity tracked semantic relatedness. We compared the similarity between MVPs evoked during the semantic retrieval of dominant-biased homonyms and their intended unambiguous synonym (e.g., “ball”; “orb”). This analysis did not yield any reliable neural similarity effects in response to semantically related versus unrelated homonym-synonym word pairs. Further, we did not find any areas in which neural similarity continuously scaled with subjective, numerical ratings of semantic relatedness.

To further characterize the M1- and left VLPFC-predicted MVPs that we identified in left ATL, we performed follow-up analyses in the peak searchlight volume. In particular, we compared the relative similarities between the MVPs evoked during retrieval of each item's dominant-biased (e.g., sphere-ball); subordinate-biased (e.g., dance-ball); and dominant-synonym (e.g., “orb”) presentations. This analysis revealed that the synonym MVPs were more similar to the dominant-biased patterns (mean $r = 0.02$) than they were to the subordinate-biased patterns (mean $r = 0.001$), $t(12) = 2.07$, $p = 0.06$. We also checked whether M1 or left VLPFC activity could predict a synonym's relative MVP similarity match to the dominant-biased homonym presentation, versus its similarity to the subordinate-biased presentation. We observed a positive relationship between left VLPFC response and meaning match, such that the synonym pattern's relative similarity to the dominant-biased versus subordinate-biased presentation is predicted by increases in left VLPFC response, $t(12) = 2.63$, $p = 0.02$. That is, when left VLPFC response increases during the subordinate-biased word presentation, its resemblance to the synonym pattern

decreases, relative to the similarity between the dominant and synonym MVPs. In contrast, M1 did not reliably predict increases in a synonym's match to the dominant versus subordinate-biased word presentation, $t(12) = 0.68$, $p = 0.51$.

These post-hoc findings in left ATL suggest that the neural patterns observed here might encode abstract, conceptual information about word meanings. Alternatively, it is possible that our neural similarity effects in left ATL could reflect the activation of lexical representations that serve as an interface between word form and meaning. With the current data and paradigm, we are unable to determine whether the MVPs that we identified in left ATL represent lexical versus conceptual information (or some combination of the two). Our interpretations of the effects in this region are limited, because although we can predict within-word neural similarity using two parametric, item-level measures (i.e., M1 and left VLPFC response) which have strong theoretical and empirical support for predicting lexical-semantic competition (cf. Rodd, 2005; Twilley & Dixon, 2000), we are nevertheless unable to describe the dimensions that govern the observed similarities. Future research will benefit from more extensively characterizing the nature of the representational similarity space evoked by lexical stimuli in left ATL.

To more conclusively determine whether left ATL activity reflects the co-activation of competing word meanings, future analyses should interrogate neural patterns evoked by additional noun-noun homonyms, and several synonyms for both dominant and subordinate homonym meanings. Additionally, more elaborate and in-depth behavioral measures of stimulus processing during sentence comprehension and semantic are necessary to make any strong claims about the extent to which disambiguating linguistic contexts might influence the resulting neural patterns. The present study is the first step in applying a combination of behavioral and fMRI multivariate analysis techniques to advance our understanding of how people interpret ambiguous linguistic input (see also Danelli et al., 2015). The current work demonstrates the promise and utility of this approach.

5.2.2. Left VLPFC activity negatively predicts within-word neural similarity in left anterior temporal lobe

The meaning frequency effects in left ATL suggest that the dominant meaning of polarized homonym words might always be retrieved, regardless of context. But does biasing context have any effect on the activation of the dominant meaning? To address this question, we tested whether BOLD response in left VLPFC tracks decreases in neural similarity between the activation patterns evoked by context-appropriate and context-inappropriate homonym meanings. This analysis revealed that when left VLPFC response increases during the subordinate meaning retrieval, within-word neural similarity decreases in left ATL. We suggest that the reductions in neural similarity reflect the task-driven expression of the subordinate, contextually appropriate word meaning, and its distinction from the initial, contextually inappropriate dominant meaning, thereby increasing the dissimilarity between their corresponding neural patterns. When a comprehender must resolve the interference caused by alternative meanings of a single word form, left VLPFC may act as a top-down modulatory signal to bias neural patterns toward the contextually appropriate representation.

Empirical support for this proposal comes from both our own data in the same set of subjects, and from numerous other studies. In the present study, we demonstrated that left VLPFC response is associated with the reinterpretation of homonym meanings, in which a subordinate meaning must be selected over an initially activated dominant meaning. Moreover, during the Stroop conflict task, responses here increased during conflict trials, during which distracting information (i.e., incongruent color names) must be

ignored. Further, evidence from converging methods, including patient lesion data, TMS, and fMRI demonstrate that this region is activated during, or is necessary for, selecting contextually-appropriate meanings of ambiguous words (Bedny, Hulbert, & Thompson-Schill, 2007; Bedny, McGill, & Thompson-Schill, 2008; Ihara et al., 2015; Rodd, 2005; Rodd et al., 2012; Thompson-Schill et al., 2005); completing sentences with multiple alternative responses (Robinson, Shallice, & Cipolotti, 2005); generating verbs with many semantic competitors (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997); and overriding misinterpretations of syntactically ambiguous sentences (January et al., 2009; Rodd et al., 2010).

The linear effect of left VLPFC response magnitude on neural similarity suggests that multiple homonym meanings compete for activation during the semantic retrieval of a single meaning, and that left VLPFC tracks the resolution of this conflict. This result is compatible with a handful of other studies that have reported a relationship between left VLPFC activity and dissimilarity between MVPs evoked by competing stimuli elsewhere in the brain. In a recent study by Hindy et al. (2015), in early visual cortex, the neural dissimilarity between MVPs evoked by two incompatible states of the same object (e.g., a cracked versus intact egg) was predicted by increased left VLPFC response during the presentation of the object in its second state.

Likewise, MVPA studies in the domain of episodic memory, recent studies have found that recruitment of frontal cortex during the encoding (Kuhl et al., 2012) and the retrieval (Wimber et al., 2015) of updated memories predicts decreased competition from earlier memories. One interesting possibility is that episodic interference from older memories may have played a role in the present study as well. In our paradigm, dominant meanings were presented in the first half of the experiment, followed by the subordinate meanings in the second half. Perhaps subjects experienced episodic interference from the memory event of comprehending and retrieving the dominant meaning earlier in the experiment. However, it is unclear how various sources of potential interference (e.g., episodic or semantic) might interact and influence lexical ambiguity resolution. This open and interesting question warrants further study.

Taken together with our findings, we propose that left VLPFC serves as a domain-general, top-down control signal that suppresses competition between co-activated neural representations, and that the outcome of this modulatory role can be identified in the dissimilarity between neural patterns evoked in posterior cortical areas. However, although the pattern-predicted increase in left VLPFC response was reliable across subjects, it was not robust at the whole-brain level. Rather, the relationship between left VLPFC response and left ATL neural similarity was identified through the fROI-based analyses, in which we limited our analyses to the fluctuations in BOLD response in subject-specific, Stroop-conflict sensitive regions of left VLPFC. Why did this relationship fail to emerge at the whole-brain level? One possibility is that there are individual differences in the extent to which left VLPFC is recruited while subjects retrieve a context-appropriate homonym meaning. In fact, previous fMRI studies on lexical ambiguity resolution have found that frontal recruitment during the retrieval of subordinate meanings can be predicted by individual differences in reading span (Mason & Just, 2007) and behavioral performance during a semantic interference task (Hoening & Scheef, 2009). Additional research is necessary to determine the subject-specific variables associated with pattern-predicted activity in left VLPFC.

Although the current study focused on the role of left VLPFC, other studies indicate that additional brain regions also participate in cognitive control processes (e.g., right prefrontal cortex and anterior cingulate cortex (ACC)), particularly when an overt

response is required. For instance, along with left VLPFC, responses in ACC and right PFC increase during judgements of homonym words (Bedny et al., 2008; Chan et al., 2004; Hoenig & Scheef, 2009) and during incongruent trials of the Stroop task (MacLeod & MacDonald, 2000). In contrast, BOLD response in ACC and right PFC was not modulated by sentence condition in our whole-brain analysis. The lack of reliable activity in these regions has also been observed in other fMRI studies that, similar to our experimental paradigm, measured BOLD response during passive comprehension of homonyms embedded in sentence contexts (e.g., Rodd, 2005; Vitello et al., 2014; Zemleni et al., 2007). This differential response profile suggests that the recruitment of brain regions implicated in cognitive control processes depends on the specific task demands (Milham et al., 2001).

5.3. Conclusions

The representation of multiple lexical-semantic representations of the same homonym word across contexts, and how these representations might compete for activation, has not been extensively studied. The data reported here suggest that not only do ambiguous word meanings compete for selection in left ATL, but also that the extent of their competition is driven by both bottom-up features (frequency-based form-to-meaning associations) and top-down neural signals (left VLPFC response magnitude). We present the first step in identifying the representational mechanisms that given rise to successful resolution of semantic ambiguity.

Acknowledgements

This research was funded by an NIH Award to Sharon L. Thompson-Schill (ROI DC009209) and an NSF graduate research fellowship to Elizabeth Musz.

References

- Bedny, M., Hulbert, J. C., & Thompson-Schill, S. L. (2007). Understanding words in context: the role of Broca's area in word comprehension. *Brain Research*, 1146, 101–114. <http://dx.doi.org/10.1016/j.brainres.2006.10.012>.
- Bedny, M., McGill, M., & Thompson-Schill, S. L. (2008). Semantic adaptation and competition during word comprehension. *Cerebral Cortex*, 18(11), 2574–2585. <http://dx.doi.org/10.1093/cercor/bhn018>.
- Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J. M., & Lambon Ralph, M. A. (2010). The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cerebral Cortex*, 20(11), 2728–2738. <http://dx.doi.org/10.1093/cercor/bhq019>.
- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977–990. <http://doi.10.3758/BRM.41.4.977>.
- Chan, A. H., Liu, H. L., Yip, V., Fox, P. T., Gao, J. H., & Tan, L. H. (2004). Neural systems for word meaning modulated by semantic ambiguity. *NeuroImage*, 22, 1128–1133. <http://dx.doi.org/10.1016/j.neuroimage.2004.02.034>.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3), 162–173.
- Danelli, L., Marelli, M., Berlinger, M., Tettamanti, M., Sberna, M., Paulsen, E., & Luzzatti, E. (2015). Framing effects reveal discrete lexical-semantic and sublexical procedures in reading: An fMRI study. *Frontiers in Psychology*, 6, 1328. <http://dx.doi.org/10.3389/fpsyg.2015.01328>.
- Detre, G., Polyn, S. M., Moore, C., Natu, V., Singer, B., Cohen, J., ... & Norman, K. A. (2006, June). The multi-voxel pattern analysis (MVPA) toolbox. In *Poster presented at the Annual Meeting of the Organization for Human Brain Mapping (Florence, Italy)*.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27, 429–446.
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, 25(4), 1325–1335. <http://dx.doi.org/10.1016/j.neuroimage.2004.12.034>.
- Fedorenko, E., Hsieh, P.-J., Nieto-Castanon, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: defining nouns functionally in individual subjects. *Journal of Neurophysiology*, 104(2), 1177–1194. <http://dx.doi.org/10.1152/jn.00032.2010>.
- Gorfein, D. S. (Ed.). (2001). *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity*. Washington, DC: American Psychological Association.
- Hindy, N. C., Altmann, G. T. M., Kalenik, E., & Thompson-Schill, S. L. (2012). The effect of object state-changes on event processing: Do objects compete with themselves? *Journal of Neuroscience*, 32(17), 5795–5803. <http://dx.doi.org/10.1523/JNEUROSCI.6294-11.2012>.
- Hindy, N. C., Solomon, S. H., Altmann, G. T. M., & Thompson-Schill, S. L. (2015). A cortical network for the encoding of object change. *Cerebral Cortex*, 25(4), 884–894. <http://dx.doi.org/10.1093/cercor/bht275>.
- Hoenig, K., & Scheef, L. (2009). Neural correlates of semantic ambiguity processing during context verification. *NeuroImage*, 45(3), 1009–1019. <http://dx.doi.org/10.1016/j.neuroimage.2008.12.044>.
- Huetig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, 15(8), 985–1018. <http://dx.doi.org/10.1080/13506280601130875>.
- Ihara, A. S., Mimura, T., Soshi, T., Yorifuji, S., Hirata, M., Goto, T., ... Fujimaki, N. (2015). Facilitated lexical ambiguity processing by transcranial direct current stimulation over the left inferior frontal cortex. *Journal of Cognitive Neuroscience*, 27(1), 26–34. http://dx.doi.org/10.1162/jocn_a.00703.
- January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, 21(12), 2434–2444.
- Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8(5), 679–685. <http://dx.doi.org/10.1038/nn1444>.
- Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103(10), 3863–3868.
- Kuhl, B. A., Bainbridge, W. A., & Chun, M. M. (2012). Neural reactivation reveals mechanisms for updating memory. *Journal of Neuroscience*, 32(10), 3453–3461. <http://dx.doi.org/10.1523/JNEUROSCI.5846-11.2012>.
- Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences*, 108(14), 5903–5908. <http://dx.doi.org/10.1073/pnas.1016939108>.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163. <http://dx.doi.org/10.1037/0033-2909.109.2.163>.
- MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, 4(10), 383–391. [http://dx.doi.org/10.1016/S1364-6613\(00\)01530-8](http://dx.doi.org/10.1016/S1364-6613(00)01530-8).
- Mason, R. A., & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research*, 1146, 115–127. <http://dx.doi.org/10.1016/j.brainres.2007.02.076>.
- Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., & Kramer, A. F. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research*, 12(3), 467–473. [http://dx.doi.org/10.1016/S0926-6410\(01\)00076-3](http://dx.doi.org/10.1016/S0926-6410(01)00076-3).
- Musz, E., & Thompson-Schill, S. L. (2015). Semantic variability predicts neural variability of object concepts. *Neuropsychologia*, 76, 41–51. <http://dx.doi.org/10.1016/j.neuropsychologia.2014.11.029>.
- Pacht, J. M., & Rayner, K. (1993). The processing of homophonic homographs during reading: Evidence from eye movement studies. *Journal of Psycholinguistic Research*, 22, 252–271.
- 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. <http://dx.doi.org/10.1038/nrn2277>.
- Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010). Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Current Biology*, 20(10), 964–968. <http://dx.doi.org/10.1016/j.cub.2010.03.070>.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Rayner, K., Pacht, J. M., & Duffy, S. A. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language*, 33, 527–544.
- Reddy, L., Kanwisher, N. G., & VanRullen, R. (2009). Attention and biased competition in multi-voxel object representations. *Proceedings of the National Academy of Sciences*, 106(50), 21447–21452.
- Ritche, M., Wing, E. A., LaBar, K. S., & Cabeza, R. (2013). Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cerebral Cortex*, 23(12), 2818–2828. <http://dx.doi.org/10.1093/cercor/bhs258>.
- Robinson, G., Shallice, T., & Cipolletti, L. (2005). A failure of high level verbal response selection in progressive dynamic aphasia. *Cognitive Neuropsychology*, 22(6), 661–694. <http://dx.doi.org/10.1080/02643290442000239>.
- Rodd, J. M. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, 15(8), 1261–1269. <http://dx.doi.org/10.1093/cercor/bhi009>.
- Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2012). Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. *Cerebral Cortex*, 22(8), 1761–1773. <http://dx.doi.org/10.1093/cercor/bhr252>.
- Sereno, S. C., O'Donnell, P. J., & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate-bias effect. *Journal of Experimental Psychology: Human Perception and Performance*, 32(2), 335.

- Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J. A., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex*, *19*(7), 1493–1503. <http://dx.doi.org/10.1093/cercor/bhn187>.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643.
- Swaab, T., Brown, C., & Hagoort, P. (2003). Understanding words in sentence contexts: The time course of ambiguity resolution. *Brain and Language*, *86*(2), 326–343. [http://dx.doi.org/10.1016/S0093-934X\(02\)00547-3](http://dx.doi.org/10.1016/S0093-934X(02)00547-3).
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645–659.
- Thompson-Schill, S. L., Bedny, M., & Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Current Opinion in Neurobiology*, *15*(2), 219–224. <http://dx.doi.org/10.1016/j.conb.2005.03.006>.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792–14797.
- Twilley, L. C., & Dixon, P. (2000). Meaning resolution processes for words: A parallel independent model. *Psychonomic Bulletin & Review*, *7*(1), 49–82.
- Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition*, *22*(1), 111–126.
- Visser, M., Jefferies, E., & Ralph, M. L. (2010). Semantic processing in the anterior temporal lobes: A meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, *22*(6), 1083–1094.
- Visser, M., & Ralph, M. L. (2011). Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *Journal of Cognitive Neuroscience*, *23*(10), 3121–3131.
- Vitello, S., Warren, J. E., Devlin, J. T., & Rodd, J. M. (2014). Roles of frontal and temporal regions in reinterpreting semantically ambiguous sentences. *Frontiers in Human Neuroscience*, *8*. <http://dx.doi.org/10.3389/fnhum.2014.00530>.
- Whitney, C., Jefferies, E., & Kircher, T. (2011). Heterogeneity of the left temporal lobe in semantic representation and control: priming multiple versus single meanings of ambiguous words. *Cerebral Cortex*, *21*(4), 831–844. <http://dx.doi.org/10.1093/cercor/bhq148>.
- Wiley, J., & Rayner, K. (2000). Effects of titles on the processing of text and lexically ambiguous words: Evidence from eye movements. *Memory & Cognition*, *28*(6), 1011–1021.
- Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., & Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nature Neuroscience*, *18*(4), 582–589. <http://dx.doi.org/10.1038/nn.3973>.
- Wing, E. A., Ritchey, M., & Cabeza, R. (2015). Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. *Journal of Cognitive Neuroscience*, *27*(4), 679–691. http://dx.doi.org/10.1162/jocn_a_00740.
- Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited—Again. *Neuroimage*, *2*, 173–181.
- Zemleni, M.-Z., Renken, R., Hoeks, J. C. J., Hoogduin, J. M., & Stowe, L. A. (2007). Semantic ambiguity processing in sentence context: Evidence from event-related fMRI. *NeuroImage*, *34*(3), 1270–1279. <http://dx.doi.org/10.1016/j.neuroimage.2006.09.048>.