RESEARCH ARTICLE

WILEY

From letters to composed concepts: A magnetoencephalography study of reading

Graham Flick[1,2] 🔘 | Osama Abdullah[2] | Liina Pylkkänen[1,2,3]

[1]Department of Psychology, New York University, New York, New York

[2]NYUAD Research Institute, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

[3]Department of Linguistics, New York University, New York, New York

Correspondence
Graham Flick, Department of Psychology, New York University, New York, NY.
Email: graham.flick@nyu.edu

Abstract

Language comprehension requires the recognition of individual words and the combination of their meanings to yield complex concepts or interpretations. This combinatory process often requires the insertion of unstated semantic material between words, based on thematic or feature knowledge. For example, the phrase *horse barn* is not interpreted as a blend of a horse and a barn, but specifically a barn where horses are kept. Previous neuroscientific evidence suggests that left posterior and anterior temporal cortex underpin thematic and feature-based concept knowledge, respectively, but much remains unclear about how these areas contribute to combinatory language processing. Using magnetoencephalography, we contrasted source-localized responses to modifier-noun phrases involving thematic relations versus feature modifications, while also examining how lower-level orthographic processing fed composition. Participants completed three procedures examining responses to letter-strings, adjective-noun phrases, and noun–noun combinations that varied the semantic relations between words. We found that sections of the left anterior temporal lobe, posterior temporal lobe, and cortex surrounding the angular gyrus were all engaged in the minimal composition of adjective-noun phrases, a more distributed network than in most prior studies of minimal composition. Of these regions, only the left posterior temporal lobe was additionally sensitive to implicit thematic relations between composing words, suggesting that it houses a specialized relational processing component in a wider composition network. We additionally identified a left occipitotemporal progression from orthographic to lexical processing, feeding ventral anterior areas engaged in the combination of word meanings. Finally, by examining source signal leakage, we characterized the degree to which these responses could be distinguished from one another using source estimation.

KEYWORDS
angular gyrus, anterior temporal lobe, composition, magnetoencephalography, posterior temporal lobe, semantics

1 | INTRODUCTION

Successful language comprehension relies on the recognition of familiar words and the combination of their meanings to yield complex concepts or interpretations (i.e., the meaning of a phrase or sentence). An intriguing part of this process is that composition often requires the specification of covert semantic relations between words to yield their full interpretation. For example, with little effort most English speakers will understand the phrase *horse barn* as a barn in which horses are kept, rather than a vague blend of a horse and a barn or a barn that looks like a horse. Similarly, a *trophy cabinet* is a cabinet in which trophies are kept, paralleling *horse barn* in its implicit semantic relation, while a *metal cabinet* differs from the two, likely understood as a cabinet made of metal.

Depending on the words being combined, identifying these relations relies on different types of semantic knowledge in memory, including feature knowledge of the constituent concepts (e.g., size or color) and thematic knowledge of how they interact (Estes, 2003; Wisniewski, 1996; Wisniewski & Love, 1998). While behavioral and neuroscientific investigations have demonstrated that individual words activate these aspects of conceptual knowledge in memory (e.g., Kalénine, Mirman, Middleton, & Buxbaum, 2012; Mirman & Graziano, 2012a, 2012b), comparatively little is known about how we identify and process the different semantic relations between words when they are needed to interpret multi-word concepts. Here, we used magnetoencephalography (MEG) to ask where and when the use of this knowledge may be supported by neural activity during the comprehension of two-word phrases, beginning at early stages of individual visual word recognition through to the integration of meanings to build composed concepts.

1.1 | Adjectives, nouns, and semantic hubs

Noun–noun combinations are a common context in which comprehenders must insert unstated semantic relations between words to complete the composed meaning (Gagné & Shoben, 1997; Murphy, 1990). Previous behavioral work on these combinations has identified two ways that the constituent nouns often relate to one another (Wisniewski, 1996; Wisniewski & Love, 1998): (a) The transfer of a feature or attribute from the modifier to the head noun (attributive interpretations: *robin snake* is a red-bellied snake, where red-belly is transferred from robin) and (b) the insertion of an implicit thematic relation between words (relational interpretations: *robin snake* is a snake that hunts robins). A recent functional magnetic resonance imaging (fMRI) study revealed separable impacts of the two interpretation types, implicating bilateral sections of temporoparietal cortex (TPC) in the processing of relation-based compounds and the left anterior temporal lobe (ATL) in the processing of attributive or feature-based compounds (Boylan, Trueswell, & Thompson-Schill, 2017). On the basis of anatomical connectivity and patterns of task-related activations, both of these regions have been suggested to house and/or function as so-called semantic hub areas (Lambon Ralph,

Jefferies, Patterson, & Rogers, 2017; Patterson, Nestor, & Rogers, 2007; Schwartz et al., 2011); a characterization reinforced by their position along gradients of resting state functional connectivity (Margulies et al., 2016). Extant accounts differ, however, in their proposals regarding what types of semantic processing are housed in one or both of these regions.

The predictions made by so-called dual hub models of semantic knowledge (de Zubicaray, Hansen, & McMahon, 2013; Schwartz et al., 2011) are consistent with the pattern found in the results of Boylan et al. (2017). These accounts propose that left posterior temporal and/or TPC is primarily involved in the processing of thematic information, while left anterior temporal cortex is primarily involved in processing of feature or taxonomic information (see also Mirman & Graziano, 2012a, 2012b, for supporting findings). Although conflicting results exist (Mirman, Landrigan, & Britt, 2017), such accounts are bolstered by the findings that TPC and neighboring sections of the left posterior temporal lobe (PTL), particularly the posterior middle and superior temporal gyri (pMTG and pSTG), are involved in the processing of word meanings believed to rely, to a greater degree, on thematic or relational knowledge (e.g., verbs relative to nouns; Bedny, Caramazza, Grossman, Pascual-Leone, & Saxe, 2008; Bedny, Dravida, & Saxe, 2014; Bedny & Thompson-Schill, 2006; Davis, Meunier, & Marslen-Wilson, 2004; Kable, Kan, Wilson, Thompson-Schill, & Chatterjee, 2005; Kable, Lease-Spellmeyer, & Chatterjee, 2002; Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995; Yu, Bi, Han, Zhu, & Law, 2012; Yu, Law, Han, Zhu, & Bi, 2011), as well as tool usage (Weisberg, Van Turennout, & Martin, 2007), and action knowledge (Kalénine & Buxbaum, 2016; Kemmerer, Rudrauf, Manzel, & Tranel, 2012; see also Kalénine et al., 2009). The left pMTG and AG have also been found to show increasing activation in the production of sentences containing verbs with increasing argument complexity (Takashima, Konopka, Meyer, Hagoort, & Weber, 2020), reflecting greater complexity in relational structure. Speaking more directly to this issue, Williams, Reddigari, and Pylkkänen (2017) implicated left superior temporal and inferior parietal activity in the processing of relationality independent of closely related factors such as syntactic category and eventivity, suggesting that it is the relevance of thematic or relational knowledge rather than, for example, verb versus noun status, that engages activity in these areas. Complementary findings have highlighted the importance of the left ATL to feature knowledge and identification of objects (Baron, Thompson-Schill, Weber, & Osherson, 2010; Clarke, Taylor, & Tyler, 2011; Coutanche & Thompson-Schill, 2015; Moss, Rodd, Stamatakis, Bright, & Tyler, 2005; Tyler et al., 2004), congruent with the expected properties of a taxonomic or feature-oriented semantic hub.

Alternatively, the controlled semantic cognition (CSC) framework (Jefferies, Thompson, Cornelissen, & Smallwood, 2020; Lambon Ralph et al., 2017) posits that the left ATL functions as the primary semantic hub and its interaction with other areas underpins both thematic and taxonomic knowledge. Based on patterns of connectivity (Davey et al., 2016) and the results of disruption (Hallam et al., 2016; Hallam et al., 2018; Whitney, Kirk, O'Sullivan, Lambon Ralph, & Jefferies, 2011), as well as responses to semantic retrieval and control tasks

(Davey et al., 2015; Teige et al., 2018, 2019), the CSC framework additionally proposes a finer distinction regarding the contributions of left posterior temporal and temporoparietal regions to semantic cognition. The angular gyrus (AG) is argued to support automatic semantic retrieval (i.e., straightforward semantic associations) while the pMTG is proposed to function as part of the brain's semantic control network, supporting the flexible retrieval of contextually relevant, but nondominant aspects of knowledge from memory (Jefferies et al., 2020). The finding of activation in this posterior temporal lobe region in response to verbs relative to nouns (as well as increased responses to actions) is then postulated to be due to greater semantic control demands associated with retrieving these types of meanings from memory (Jefferies et al., 2020; Thompson et al., 2017). The proposed pMTG and AG dissociation is further supported by the results of multiple meta-analyses, which found that the left AG is reliably activated by automatic semantic tasks (Humphreys & Lambon Ralph, 2015), and the left pMTG by demands related to semantic control (Jackson, 2020; Noonan, Jefferies, Visser, & Lambon Ralph, 2013).

Assuming appropriately balanced stimuli sets, with respect to semantic control demands, the results of Boylan et al. appear to suggest that dual-hub frameworks best capture how conceptual knowledge of individual word meanings are utilized to form a coherent multi-word concept in comprehension, with thematic linkages relying on the TPC, specifically the AG, and feature-based linkages relying on the left ATL. However, the stimulus set of Boylan et al. contained many novel or otherwise unfamiliar noun–noun combinations (e.g., *sponge memory*) and the nature of the procedure provided participants with time to explicitly consider each phrase's meaning. This may have caused enhanced conscious deliberation regarding the interpretation of each combination, raising the possibility that the observed TPC and ATL responses were tied to those particular conditions. In contrast, in everyday language use we frequently encounter familiar combinations of words whose complete interpretations rely on either thematic links (e.g., *horse barn*) or the modification of a single feature of a word's meaning. A common example of the latter is the combination of color-denoting adjectives with nouns that denote relatively simple concrete objects (e.g., *brown barn*).

To the best of our knowledge, there has not yet been a direct comparison of how simple and familiar adjective-noun and noun–noun combinations, like those above, differentially tax neural processing in the proposed hub and control regions, keeping the task, paradigm, and participants constant across the conditions. In such a contrast, strict dual hub accounts would predict, as in the Boylan et al. results, that demands related to thematic linkages in relational noun–noun compounds should engage the AG and/or posterior temporal regions, while feature-based adjective-noun modifications should exclusively modulate the left ATL's responses. Alternatively, the left ATL may show both of these sensitivities, without corresponding modulation in the posterior temporal/temporoparietal regions, consistent with models that posit a single conceptual hub underpinning both feature and thematic knowledge.

1.2 | A ventral anterior hub supporting composition

Another issue relevant to neuroanatomical accounts of semantic knowledge, and the way this knowledge is used in language comprehension, concerns how neural responses related to the sensory processing of a word feed into regions proposed to underpin access to knowledge of that word's meaning. This is a particularly pertinent question in visual reading, as the proposed ventral location of the ATL hub (Chen et al., 2016) lays in the vicinity of occipital and temporal lobe areas known to support visual word recognition (Cohen et al., 2002; Dehaene et al., 2010; Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999; Taylor, Davis, & Rastle, 2019; Woolnough et al., 2020). Previous MEG studies of reading, which provide the temporal resolution to tease apart distinct processing stages in word recognition (Pylkkänen, Stringfellow, & Marantz, 2002; Tarkiainen et al., 1999), have identified a consistent sequence of responses along the left ventral occipital and temporal lobes within 200 ms after visual word onset, progressing from sensitivities to low-level visual properties to the discrimination of word versus symbol-string stimuli in the middle and anterior fusiform and inferior temporal gyri (Gwilliams, Lewis, & Marantz, 2016; Neophytou, Manouilidou, Stockall, & Marantz, 2018; Tarkiainen et al., 1999). This progression has been proposed as an electrophysiological manifestation of the hierarchy of visual word responses observed in fMRI studies (e.g., Vinckier et al., 2007), feeding into left temporal lobe regions sensitive to a word's phonology and meaning (Taylor et al., 2019; Wang, Deng, & Booth, 2019). It has also been demonstrated to include the generators of well characterized MEG evoked components, including the M130 and M170, proposed to underpin orthographic (M130) and early lexical (M170) processing (Gwilliams et al., 2016). Moreover, intracranial recordings have demonstrated that a first pass of activity to the ventral ATL enables the discrimination of a word's semantic class (object vs. animal) by as early as 130 ms after onset (Chan et al., 2011), and that the latency of the left middle fusiform's discrimination between words and nonwords is sensitive to lexical frequency (Woolnough et al., 2020), suggesting that activity in this region indexes a search to match visual letter-strings to a meaning in memory.

Altogether, the evidence to date suggests the existence of mid and/or anterior inferior temporal lobe activity linking visual letter-string inputs to word meaning. Here, we went a step further and asked whether this same area, identified via a functional localizer, also contributes to the combination of word meanings in two-word phrases. This was motivated by two factors. First, the apparent proximity between the previous effects related to word recognition and the proposed ventral location of an ATL semantic hub (Chen et al., 2016). We reasoned that if these are one and the same region, and assuming that a semantic hub should respond with greater magnitude to combinations of word meanings as compared to single meanings, then demands related to composition should also lead to increased responses in this area. Moreover, by comparing this region's responses to different types of composition (i.e., thematic vs. feature-based) we could test the predictions of dual and single hub accounts (see above). Second, previous MEG studies of

adjective-noun composition (Bemis & Pylkkänen, 2011, 2013) have demonstrated that sections of the lateral and inferior ATL, with the specific location varying across studies, show increased activation in response to composition relative to noncombinatory stimuli (e.g., single words or word lists), modulated by conceptual-semantic properties of the composing words (Westerlund & Pylkkänen, 2014; Zhang & Pylkkänen, 2015). These effects were found at approximately the same timing that Gwilliams et al. (2016) reported that the left anterior fusiform discriminated between letters and symbol-strings (150–200 ms after onset). This too suggests that there may be a single generator of both responses, related to recognition of individual word meanings and the combination of meanings to form a phrase. Presently, this remains an open question, as there has not been an investigation that jointly localizes the temporally resolved ventral stream responses to letter and word stimuli, as well as the left ATL's responses to word combinations.

1.3 | The present study

Here, we conducted a series of MEG studies in a common sample of participants to address the issues outlined above. We began by attempting to replicate the results of previous MEG studies of visual word recognition (Gwilliams et al., 2016) and phrasal composition (Bemis & Pylkkänen, 2011). Using a letter-string response localizer developed by Gwilliams et al., we expected to observe a posterior-to-anterior progression of responses supporting visual letter-string processing, culminating in a ventral ATL site that discriminated words and symbol strings. We also expected that the phrasal composition paradigm of Bemis and Pylkkänen (2011), contrasting responses to two-word phrases, two-word lists, and single words, would reveal composition-related activity in the left ATL, and perhaps other lateral left hemisphere areas. We then built on these findings by asking the following questions: First, is there evidence of spatial overlap between the visual word recognition and combinatory responses in the ventral ATL, between 150 and 300 ms after word onset? If so, this would suggest a single node in the visual word recognition pathway contributes not only to the mapping of visual letter-strings to word meanings, but also the combination of word meanings in visual reading. Second, returning to the topic of semantic relations, in a final experiment we asked whether phrases involving thematic relations versus featural modifications differentially engaged putative semantic hub areas in the left ATL, left posterior temporal lobe, and left AG. This allowed us to test the prediction of dual hub models that thematic relations should exclusively modulate left posterior temporal or inferior parietal areas, while featural modifications should only modulate responses in the left ATL.

2 | MATERIALS AND METHODS

2.1 | Participants

Twenty-eight right-handed native English speakers (20 females, 8 males, mean age = 28.14 years, sd = 10.25 years) took part in the

study. All participants had healthy or corrected-to-healthy vision and healthy hearing. Twenty-one of the participants also took part in an MRI experiment that included collection of a high-resolution anatomical MRI, which was used in the source estimation of MEG responses.

2.2 | Stimuli and experimental design

The complete experimental design consisted of three separate MEG experiments to probe the processing of (a) visual letter-strings and words, (b) adjective-noun composition; and (c) noun–noun composition. Figure 1 displays example trials from each experiment. More details are provided in each of the following sections.

2.2.1 | Letter-string response localizer

First, to characterize responses to visual letter-strings, we adopted the localizer of Gwilliams et al. (2016, see original paper for full details), developed from the work of Tarkiainen et al. (1999). Briefly, participants passively viewed four types of visual stimuli: An individual letter (e.g., A), a combination of four letters that formed a familiar, disyllabic word (e.g., ATOM), a single shape symbol length-matched to the one-letter stimuli (e.g., a single square), and four shapes length matched to the four-letter word stimuli (e.g., a square, circle, triangle, and diamond; see Figures 1 and 2 for example stimuli). Letter and word stimuli were embedded in two levels of visual noise, defined as zero-mean Gaussian distributions with variances of 0.0234 (low) and 1.5 (high). This facilitated the isolation of neural responses sensitive to low-level visual properties of letter-strings (e.g., high vs. low visual noise across letters and words) and differences in stimulus type (e.g., letters and words vs. length-matched symbol strings). Notably, the four-letter stimuli in this localizer were pronounceable words, rather than pseudowords, meaning that their contrast with length-matched symbol strings would reveal responses that may be plausibly related to orthography, phonology, or word meaning.

2.2.2 | Experiment 1: Adjective-noun composition

Second, to probe responses related to adjective-noun composition, as in simple color-noun phrases (e.g., red + boat), we performed a truncated replication of the MEG study conducted by Bemis and Pylkkänen (2011), using all of the original stimuli and procedures (see original paper for full details). Participants performed two tasks in a blocked fashion, with each block made up of two- and one-word trials. In "Composition" blocks, two-word trials consisted of a color-denoting adjective (e.g., red) followed by one of twenty-five concrete nouns (e.g., boat). On one-word trials the adjective was replaced by a length-matched, nonpronounceable, consonant string. Following presentation of these stimuli, an image of a colored line drawing appeared on screen and participants indicated whether it matched or mismatched all of the words on that trial (i.e., was it a red boat?).
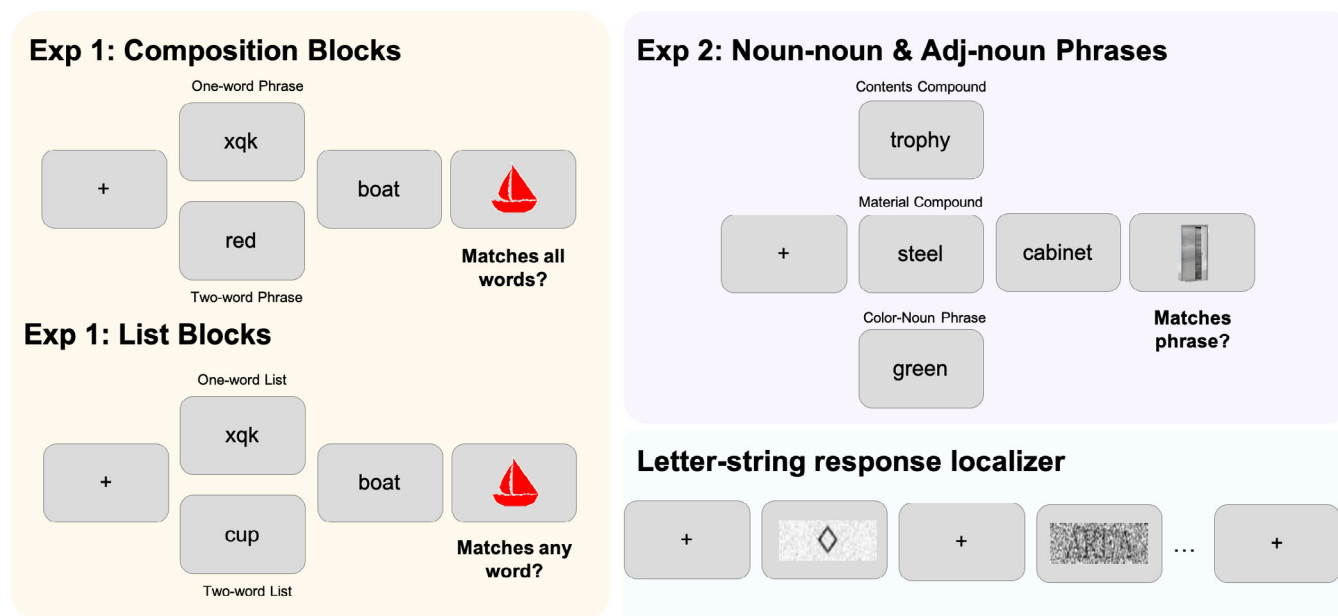
FIGURE 1 Trial structures for each experimental procedure. Left: Experiment 1 replicated the procedure of Bemis and Pylkkänen (2011) and contained two types of blocks. In Composition Blocks, participants read two-word phrases (red boat) or one-word phrases, wherein the modifier was replaced with an unpronounceable consonant string (xqk boat). Their task was to indicate if a subsequent picture matched all of the words they read on that trial. In List Blocks (bottom), participants read lists of two nouns (cup, boat) or one-word lists (xqk, boat), and then indicated if a subsequent picture matched any of the words on that trial. Top-right: In Experiment 2, designed to parallel Experiment 1, participants read three types of phrasal stimuli: Contents Compounds (modifier specifies the function/contents of the head noun, considered a thematic relationship; trophy cabinet), Material Compounds (modifier specifies the material of the head noun, considered a nonthematic relationship; metal cabinet), or adjective-noun phrases in which the adjective denoted a color (green cabinet; nonthematic). They were then asked to judge whether a following picture matched the meaning of the phrase they read on each trial. Bottom-right: Participants also completed a letter-string response localizer task (see Gwilliams et al., 2016 for full details), which involved passive viewing of letter- and symbol-string stimuli. See the main text for complete details on all three procedures

In "List" blocks, two-word trials consisted of a pair of nouns (e.g., cup, boat), rather than a phrase, while one-word trials again replaced the first position item with a length-matched consonant string. As in Composition blocks, an image appeared following the word stimuli, but here participants were asked to indicate whether the pictured matched any of the words on that trial (i.e., was it either a cup or boat?). On one-word trials, in each task, participants only needed to consider the single noun they encountered. Each participant completed 50 trials in each condition in four alternating Composition/List blocks. This represented one half of the original procedure (which had 100 trials in each condition). Stimuli lists were pseudo-randomly generated for each participant such that no unique combination of word/nonword stimuli could be repeated more than twice, no list could consist of a repeated noun, and each condition was balanced on the overall length of the first position item. We note that in the original paradigm, Bemis and Pylkkänen (2011) repeated the same set of images in each condition. Here, instead, we pseudo-randomly selected a set of 200 total images for each participant, which were then assigned to each trial to ensure that half of each condition was followed by a match and half followed by a mismatch. In the two-word conditions, the image assignments were also appropriately balanced on the relevance of the first or second item to determining whether the correct response was match or mismatch.

2.2.3 | Experiment 2: Noun–noun composition

The third procedure was modeled after the previously described adjective-noun composition paradigm (particularly the two-word Composition trials), but contrasted responses to noun–noun combinations that differed in their relational structures between modifier and head, as well as the same set of head nouns modified by color-denoting adjectives (e.g., trophy cabinet vs. metal cabinet vs. green cabinet). The construction of the stimulus set began with the selection of fourteen container-denoting head nouns, which could be straightforwardly modified by a preceding noun specifying the material they are made of (e.g., metal cabinet) as well as a noun in a spatial/functional relationship with the head (e.g., a trophy cabinet is a cabinet that functions by containing trophies). For convenience, we refer to the former as Material modifiers (and their corresponding whole phrases as Material Compounds) and the latter as Contents modifiers (and Contents Compounds). Fourteen Content modifiers were selected, and each was paired with between 2 and 5 head nouns to yield a set of 42 items (see Supporting Information for the complete stimulus set). Fourteen Material modifiers were then selected such that each was assigned to replace one Contents modifier in all of its phrases, creating a set of 42 Material Compounds. Notably, although a subset of the phrases in both conditions were novel (e.g., shampoo

cup), the complete set in each condition was designed to include primarily familiar noun–noun combinations (see below for more details on familiarity).

Lexical and phrasal characteristics of the stimuli were extracted from the Corpus of Contemporary American English (Davies, 2009) and the English Lexicon Project (Balota et al., 2007). The selected modifiers were appropriately balanced on word length (Contents: mean = 6.00, SD = 1.82; Material: mean = 5.69, SD = 1.73), lexical frequency (Contents: mean = 13,213.98, SD = 9,592.3; Material: mean = 13,816.86, SD = 10,279.7), and mean reaction time in lexical decision tasks, as reported in the English Lexicon Project (Contents: mean = 635.53 ms, SD = 65.48 ms; Material: mean = 638.56, SD = 78.24 ms). The two sets of compounds were balanced on bigram frequency when the head noun was marked as a singular noun (Contents: mean = 41.00, SD = 81.66; Material: mean = 37.67, SD = 74.17) and transition probability from modifier to head (Contents: mean = 0.004, SD = 0.009; Material: mean = 0.004, SD = 0.008). The two sets also had comparable numbers of phrases with zero bigram frequency in the Corpus of Contemporary American English (9 items in the Material Compounds and 13 in the Contents Compounds). For context, of the 64 items in each of Boylan et al.'s (2017) relational and attributive compound sets (selected from previous studies), 26 and 37 had zero bigram counts in this corpus, respectively. We additionally conducted two stimulus norming studies on Amazon Mechanical Turk (AMT) to confirm that the sets were balanced on overall familiarity and readers' tendencies to interpret the phrase using the intended relational structure (i.e., a metal cabinet is a cabinet made of metal, rather than a cabinet that holds metal). For complete details of these norming studies, see the Supporting Information. In brief, Contents and Material compounds were rated as similarly familiar (using a Likert rating scale where 1 indicates completely unfamiliar and 7 indicates extremely familiar: Material: m = 4.750, SD = 1.021; Contents: m = 4.838, SD = 1.351) and were consistently interpreted in the intended fashion.

Once the complete noun–noun stimuli had been finalized, we also matched each modifier to one of fourteen color-denoting adjectives. It was not possible to balance the noun modifiers and color adjectives on the same properties as the two original conditions, with the color adjectives being more frequent on average and showing greater variability in their frequency (m = 92,030.74, SD = 105,749.29), shorter in length (m = 5.02, SD = 0.92), and appearing with the head nouns with lower bigram counts (m = 22.83, SD = 55.94) and transition probabilities (m = 3.6×10^{-4}, SD = 0.001). For this reason, all critical contrasts were conducted between the Material and Contents Compounds, and the adjective-noun stimuli in this procedure were only considered in post-hoc assessments. From one perspective, Material modifiers and Color modifiers are more similar to each other than the Color modifiers are to the Contents modifiers, as the former pair specifies simple physical/visual attributes of the head noun that often covary (e.g., materials such as leather or steel each have a prototypical color). Many material modifiers can also be licit in syntactic contexts that are typically characteristic of adjectives, such as following linking verbs (e.g., that looked painful vs. that looked metal vs. that looked

trophy). Moreover, both the Material and color modifiers can be adequately paraphrased as nonpredicating relations with the head noun while the Contents Compounds cannot. That is, while "a green cabinet is a cabinet that is green" is an adequate paraphrase and "a metal cabinet is a cabinet that is metal" is underspecified but still acceptable, the phrase trophy cabinet cannot be paraphrased as "a cabinet that is (a) trophy." While only an informal analysis, this contrast serves to highlight the extra thematic relation that is required in the Contents Compounds but can be acceptably dropped in the Material Compounds and color-noun phrases.

Last, in addition to the phrasal stimuli, we also presented all individual constituent words in the third procedure in isolation, for the purposes of analyses that are beyond the scope of this paper. This was implemented by splitting the experiment into alternating two-word and one-word blocks. In the former, participants saw a randomly ordered set of the adjective-noun and noun–noun stimuli, while in the latter they saw a randomly ordered set of the individual nouns and adjectives. Each modifier constituent was presented in isolation an equal number of times as it was encountered in the phrasal contexts, while head nouns were each presented once.

As in the composition trials of the adjective-noun procedure (Experiment 1), each trial was followed by an image and participants indicated whether it matched or mismatched all of the words encountered on that trial. Participants were additionally instructed that if the image contained any depiction of the words on the trial, they should respond "match." This was specified to handle the inclusion of Material and Color adjectives in one-word blocks, as these specify only a property of objects depicted in an image (i.e., is "green" in this image?). No "list tasks" were performed in the noun–noun composition procedure. One-hundred and twenty-six images that depicted the set of adjective-noun and noun-noun phrases were selected for use as task images. A list of stimulus-image pairings was pseudo-randomly generated for each participant, such that half of the trials contained a match and half of the trials contained a mismatch, in each condition, in each block.

2.3 | Procedure

All experimental procedures took place in the Magnetoencephalography and Magnetic Resonance Imaging Laboratories of New York University Abu Dhabi. Every participant provided informed consent prior to taking part in the research and all procedures were approved by the Institutional Review Board of New York University Abu Dhabi. Prior to beginning MEG procedures, each participant had their head shape, the future locations of five head position indicator coils, and the position of three fiducial landmarks (nasion, and left and right tragi), digitally recorded using a Polhemus FastSCAN system (Polhemus, Vermont, USA). Each participant completed the MEG procedures in the following order: adjective-noun composition (Experiment 1), noun–noun composition (Experiment 2), and then the letter-string response localizer. The order of the experiments was kept constant, rather than counter-balanced, so as to reduce the possibility

that participants would be inclined to interpret the two-word lists of Experiment 1 as noun–noun combinations (e.g., cup, boat) after completing Experiment 2, which contained sensible noun–noun phrases. The order of the blocks within Experiments 1 and 2 were counterbalanced across participants. Both Experiments 1 and 2 began with a brief instruction period followed by a practice session of 20 trials of each type in that experiment. The letter-string response localizer paradigm was completed exactly as described by Gwilliams et al. (2016).

Experiments 1 and 2 each used rapid serial visual presentation (RSVP) paradigms. Example trials are shown in Figure 1. Every trial in both procedures began with the presentation of a fixation cross for 300 ms followed by a blank screen for 300 ms, and then each word on that trial presented for 300 ms, with a 300 ms blank screen following its offset. Task images were presented on screen until participants pressed the button to respond. The duration of each interval between successive trials was randomly sampled from a uniform distribution consisting of discrete values of 300, 400, 500, 600, and 700 ms.

2.4 | Data collection and preprocessing

Continuous MEG data were acquired throughout all experimental procedures with a 208-channel Kanazawa Institute of Technology system (Eagle Technology, Japan) at a sampling rate of 1,000 Hz. Online high- and low-pass filters of 0.1 and 200 Hz were used during data collection. Head position indicator coils were used to record each participant's head position, relative to the MEG sensors, before and after each experimental procedure. Twenty-one of the participants also took part in an MRI session that included the acquisition of T1- and T2-weighted high-resolution anatomical MRIs on a 3-Tesla MAGNETOM Prisma scanner (Siemens Healthineers, Erlangen, Germany). The high-resolution anatomical scans along with associated field-maps were acquired and preprocessed according to the Human Connectome Project's Young Adults protocols (Glasser et al., 2013). For all relevant participants, MRI acquisition followed participation in the MEG experiment by no more than 14 days.

MEG data from each participant were first cleaned of environmental electromagnetic noise using the Continuously Adjusted Least-Squares Method (Adachi, Shimogawara, Higuchi, Haruta, & Ochiai, 2001) based on data collected at three reference channels placed away from the head. All remaining preprocessing analysis steps were completed using the MNE-Python (v. 0.20; Gramfort et al., 2013) and Eelbrain (v. 0.28; DOI: 10.5281/zenodo.1444075) packages in the Python computing environment. The data were low-pass filtered at 40 Hz and Independent Component Analysis was used to remove data patterns matching the profile of known artifacts (eye blinks, movement-related activity, and well-characterized external noise sources). The continuous MEG data were then split into epochs around the onset of critical events in each procedure. Baseline correction was applied using prestimulus intervals. Channel noise covariance matrices were estimated for each participant from the concatenation of these intervals using the automated method of Engemann and Gramfort (2015) to select the best estimator from three options: the empirical covariance, diagonal

loading, and a data-driven extension of the Ledoit–Wolf (Ledoit & Wolf, 2004) shrinkage model. We confirmed in follow-up analyses that the removal of baseline correction did not change the general pattern of results reported below. In analyses of Experiments 1 and 2, trials in each condition were averaged within each participant's data set before source estimation. In the letter-string response localizer, following the original analysis (Gwilliams et al., 2016), individual trial data were used in source estimation without averaging. For those participants with anatomical MRIs available, images were processed with the automated segmentation algorithms of the Freesurfer software suite (http://surfer.nmr.mgh.harvard.edu/) to generate a cortical surface reconstruction and corresponding parcellations for each individual (Dale, Fischl, & Sereno, 1999; Desikan et al., 2006; Fischl, Sereno, & Dale, 1999; Fischl et al., 2004). For the remaining participants, the cortical surface of Freesurfers's "fsaverage" template and the corresponding parcellations were scaled to match the head shape and location of fiducial landmarks from each participant.

MEG and MRI coordinate spaces were co-registered based on the location of fiducial landmarks and head position indicator coils. Forward models were estimated for each participant using a single layer conductance boundary element model. The L2-minimum norm method was used to estimate source-level activity in each participant's cortical surface with noise normalization to yield dynamic statistical parameter mapping (dSPM; Dale et al., 2000). The regularization signal-to-noise ratio (SNR) parameter was set to 3 for all analyses that involved averaging in sensor space prior to source estimation. For the single trial regression analyses, because sensor-level responses were not averaged in source space, precluding the attenuation of nonstimulus-locked noise that comes from averaging, a reduced SNR parameter of 2 was used. This choice of SNR was also used in previous studies that have adopted this letter-string response localizer (Neophytou et al., 2018; Stockall, Manouilidou, Gwilliams, Neophytou, & Marantz, 2019).

As has been discussed in previous work (Dale & Sereno, 1993; Gwilliams et al., 2016; Lin, Belliveau, Dale, & Hämäläinen, 2006), anatomical information concerning the geometry of each individual's cortical surface can be used to constrain the minimum norm estimate. Specifically, the orientation of sources distributed throughout the cortical surface can be specified to lie perpendicular to it. The use of this "fixed" orientation is motivated, in large part, by the known sensitivity of MEG to postsynaptic potentials in pyramidal cells, which lie perpendicular to the cortical surface (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993; Okada, Wu, & Kyuhou, 1997). Alternatively, the orientation of the source with respect to the cortical surface can be discarded ("free" orientation analysis) and, instead, only the magnitude (norm) of the vector returned. In the previous work from which our letter-string response localizer was taken, Gwilliams et al. (2016) demonstrated that the discarding of source orientation information (i.e., using free orientation estimates) obviated response dynamics that were relevant to the processing of visual word stimuli. This motivated our use of the fixed orientation option in the analysis of our data. Notably, this means that dissociations between conditions can be found as both positive and negative differences in estimated

responses. This polarity of estimated source amplitude depends on multiple factors, including location in the cortex and position relative to the sensors, and it is not uncommon for nearly perfectly mirrored patterns of responses to be found in adjacent areas of the cortical surface due to the folding (see Gwilliams et al., 2016 for further discussion). For these reasons, we did not attribute functional interpretation to the specific sign of differences or interpret, for example, greater negative estimated responses as deactivation rather than activation. Instead, we focused on the presence of the expected dissociation between conditions within estimated response magnitudes, regardless of whether they are positive or negative. In all source estimations, cortical patch statistics were used to define the orientation normal to the surface (Lin et al., 2006).

2.5 | Statistical analyses of MEG data

2.5.1 | Analysis: Letter-string response localizer

The primary goal of the letter-string response localizer was to replicate the pattern of results found by Gwilliams et al. (2016). For this reason, all tests were motivated by the previous results and, except where indicated, directly replicated the original analysis procedure. These focused on the identification of three primary response components representing a posterior-to-anterior progression along the ventral surface: (a) The Type I Noise component localized to lateral and ventral occipital cortex between 100 and 130 ms after stimulus onset, appearing as greater negative responses to high visual noise than low visual noise; (b) The Type II Noise component, localized to the left posterior fusiform gyrus and appearing as an opposing divergence between high and low visual noise stimuli approximately 100–130 ms after stimulus onset; and (c) The String Type component, localized to the mid and anterior fusiform and inferior temporal gyrus, appearing as greater response amplitudes to letter-string stimuli as compared to length-matched symbols between 130 and 200 ms.

Data from each individual trial were converted to distributed source amplitudes using each subject's noise-normalized minimum norm inverse solution. In source space, linear regression models were fit to the data at each time point across all trials. Models included the binary-coded predictors of interest (noise type, string type) and, as a nuisance variable, the elapsed number of trials. Spatiotemporal cluster tests (Maris & Oostenveld, 2007) were then performed on the regression coefficients for each predictor, across participants, using one-sample t-tests for significant deviations from zero. In all analyses, including those for Experiment 1 and 2 (below), clusters were formed using the threshold-free cluster enhancement method (TFCE; Smith & Nichols, 2009). Permutation p-values were estimated by random sign flipping of the coefficients ten thousand times and repeating the clustering procedure, then comparing the observed cluster magnitude (the sum of constituent statistics) to the resulting distribution of largest cluster magnitudes from each permutation.

Regions of interest (ROIs) were defined using the anatomical parcellations of the fsaverage cortical surface from Freesurfer

(Desikan et al., 2006; Van Essen, 2005) and again followed those used by Gwilliams et al. (2016). The tests for Noise Type I and II components were performed on activity between 100 and 180 ms after stimulus onset, in a spatiotemporal ROI encompassing the lateral occipital cortex, cuneus, lingual, pericalcarine, fusiform, middle temporal, and inferior temporal gyri. The test for the String Type component was performed in the anterior half of this ROI, between 130 and 300 ms after stimulus onset. The more constrained anterior ROI was motivated by the localization of the component by Gwilliams et al., and the present interest in isolating a ventral ATL region. The longer test window, extending to 300 ms, was motivated by the desire to identify potential overlap in time with adjective-noun composition effects (see below).

2.5.2 | Analysis: Experiment 1

The primary goals of Experiment 1 were to replicate the left ATL composition effect previously reported by Bemis and Pylkkänen (2011); (see Pylkkänen, 2020 for a review of subsequent replications), assess whether other lateral left hemisphere ROIs showed composition effects (a pattern previously found, but less often; Bemis & Pylkkänen, 2013), and assess whether ROIs identified from the letter-string response localizer also housed a sensitivity to composition. Tests were performed in left ATL, PTL, and AG ROIs, as well as functional ROIs defined from the results of the letter-string response localizer, located on the ventral surface of the left hemisphere. Composition effects were defined as a specific interaction in the 2×2 design: responses to the second word on each trial (i.e., the phrasal head) that uniquely differentiated the two-word adjective-noun phrases from the remaining three conditions. For each participant, responses in each condition were averaged in sensor-space before being converted to noise-normalized source estimates.

Spatiotemporal cluster tests (Maris & Oostenveld, 2007) were used to test for the presence of this interaction by specifying clusters be formed from the smallest magnitude of three t-values: the comparisons of two-word phrases versus two-word lists, two-word phrases versus one-word phrases, and two-word phrases versus one-word lists. All three t-tests were performed at each source and time point in the spatiotemporal test window, and the smallest of the three absolute values was assigned to each point. Clusters were then formed from these smallest, absolute-valued statistics using the TFCE method, and 10,000 permutations were used to estimate cluster p-values. Notably, due to the use of the absolute values in the first stage of the test, the resulting clusters could contain patches of cortex that showed both negative- and positive differences between conditions, causing the difference within the cluster, when averaged over sources, to be approximately zero. For this reason, cluster time-courses were extracted with MNE-python's "mean flip" method, which finds the dominant direction or orientation (positive or negative) across the timeseries of all of the sources in each label and applies a sign-flip to series from constituent sources that have the opposite polarity, thus increasing uniformity. When analyzing

the functional ROIs from the letter-string response localizer, since these contained a much smaller number of sources, activity was first averaged over sources and the cluster test procedure was applied in the temporal domain (see Maris & Oostenveld, 2007).

The left ATL ROI was defined as the anterior half of the superior, middle, and inferior temporal gyrus, as well as the temporal pole. This was designed to encompass those regions in which previous minimal composition MEG paradigms have identified the effect of interest (e.g., Bemis & Pylkkänen, 2011; Flick et al., 2018; Westerlund & Pylkkänen, 2014). A left PTL ROI was defined as the posterior half of the superior and middle temporal gyri and the anterior one-third of Brodmann Area 39. This ROI was designed to contain the pMTG site examined by Teige et al. (2019) and implicated in semantic control manipulations by the meta-analysis of Noonan et al. (2013). The left AG ROI was defined as the posterior two-thirds of Brodmann Area 39 and the whole of Brodmann Area 40, so as to capture the "automatic semantic processing site" identified in the meta-analysis of Humphreys and Lambon Ralph (2015). Primary tests were performed in three windows relative to the onset of the phrasal head: 150–300, 300–450, and 450–600 ms. These windows were motivated by the previous timing of left ATL composition responses (150–300 ms and 450–600 ms in Bemis & Pylkkänen, 2011, and Flick et al., 2018, respectively) and the inclusion of the intervening time.

2.5.3 | Analysis: Experiment 2

The goal of Experiment 2 was to determine which, if any, of the lateral left hemisphere ROIs showed a distinction in their responses to the two types of noun–noun compounds (i.e., Contents vs. Material) related to a distinction between thematic versus feature-based combinations. Two approaches were adopted in the analysis of Experiment 2. First, clusters that were found to show significant adjective-noun composition effects in Experiment 1, within each of the three lateral ROIs (the left PTL, ATL, and AG), were adopted as functional ROIs. The timeseries for each condition, from each participant, was extracted from each ROI using the mean flip method and temporal cluster-based tests were conducted on these time series. Clusters were formed from *t*-values comparing responses to the head nouns in the Material and Contents compounds, with post-hoc comparisons to the adjective-noun phrases used to aid interpretation. Second, rather than using the sub-regions in each of the ROIs, defined from the Experiment 1 clusters, we also conducted spatiotemporal tests across each ROI to address the possibility that a different set of sources in each region may show adjective-noun composition effects in Experiment 1, and compound type effects in Experiment 2. All results were corrected for the total number of comparisons, including both the temporal and spatiotemporal tests.

We additionally tested the ventral ATL functional ROIs that were identified in the letter-string response localizer, and which showed composition effects in Experiment 1, to see if they distinguished between the Contents and Material compounds (and adjective-noun phrases). This was done by extracting the timeseries in these ROIs

and then performing temporal cluster tests. We note that the presentation of the isolated words in Experiment 2 could be used to create statistical contrasts of responses to nouns in isolation and as phrasal heads (e.g., cabinet vs. trophy cabinet). However, the absence of a placeholder stimulus for the modifier (as in Experiment 1: xqk boat) effectively precludes this possibility, as the comparison would be confounded by differences in the baseline windows. For this reason, we abstained from analyzing the two- versus one-word contrast in Experiment 2, relying instead on inferences from the appropriately designed contrast in Experiment 1.

2.5.4 | Correction for multiple comparisons

Unless otherwise indicated, *p*-values in the primary tests of each dataset were corrected for multiple comparisons across ROIs and time windows using the false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995) with an adjusted significance threshold of *p* <.05. In the letter-string response localizer, this included the tests for Type I and II Noise Level responses, and the String Type response. In the analysis of Experiment 1, this included the spatiotemporal tests in the left ATL, PTL, and AG, as well as the temporal tests performed in the functional ROIs from the letter-string response localizer. In Experiment 2, this included both sets of tests conducted in the left ATL, PTL, and AG ROIs, as well as the String Type ventral ATL functional ROIs. Following the reasoning that each observed cluster was representative of its encompassing ROI and time-window (see discussion in Sassenhagen & Draschkow, 2019), the smallest *p*-value in each test window or ROI was submit to the correction procedure to determine if there was sufficient evidence to support a deviation from the null hypothesis.

2.6 | Spatial resolution and source crosstalk

The minimum norm approach to the MEG inverse problem is a linear method that can be examined and characterized based on how well it distinguishes activity at individual sources or, conversely, blurs this activity together (Hauk, Stenroos, & Treder, 2019). This blurring is often referred to as "signal leakage" and is inherent to solutions to the MEG inverse problem. One consequence of this is that the localization of a particular response pattern to a particular region may not be veridical, because estimated responses at that location could be conflated with activity from other locations. One way to examine this leakage is through the crosstalk functions (CTFs) or point spread functions (PSFs) for constituent sources, which provide complementary information. For each individual, the source estimation procedure requires specification of a forward matrix (based on anatomy and specifying the transformation from source- to sensor-space) and an estimated inverse matrix (specifying the transformation from sensor-to- source-space). CTFs and PSFs can be extracted from the product of these two matrices, referred to as a resolution matrix (see Hauk et al., 2019 an approachable introduction to spatial resolution metrics,

including the derivation of the resolution matrix, for more detailed discussions see Liu, Dale, & Belliveau, 2002; Molins, Stufflebeam, Brown, & Hämäläinen, 2008; Hauk, Wakeman, & Henson, 2011). For a particular source i, the CTF can be found in the i-th row of the resolution matrix, and captures how unit magnitude at each source could leak into the estimate for source i. Conversely, the PSF for source i describes how activity from that source leaks or spreads into the estimated activity at all other sources and can be found in the i-th column of the resolution matrix.

We examined spatial accuracy and overlap of CTFs for each of the regions identified in the analyses below. For each area, we computed the CTFs for all sources contained within it to determine if the pattern may be generated or influenced by activity elsewhere in the cortical surface. This was particularly important when examining patches of cortex that showed similarly timed effects, as they could be either independent responses or influenced by shared signal leakage. We also examined each region's peak localization error (PLE) based on the PSFs of constituent sources. For an individual source, PLE was defined as the distance between the location of the source and the absolute maximum of that source's PSF. In an ideal scenario, a source's PLE is equal to zero (i.e., the location of the maximum is the source itself). For a patch of cortex, the PLE and CTF were computed as the mean over all constituent sources.

3 | RESULTS

3.1 | Behavioral results

Accuracy data collected in Experiments 1 and 2 were examined for the purpose of confirming attendance to the experimental procedures. Accuracy was high across both experiments, indicating that participants understood and attended to the tasks. In Experiment 1, mean accuracy was 95.7% correct ($SD = 4.72\%$). All participants scored above 84% accuracy and only four below 90%. Coding of accuracy in Experiment 2 was more ambiguous due to the increased complexity of the words' depictions (i.e., steel, titanium, canvas, etc.) and the corresponding increases in the complexity and ambiguity of the image stimuli. Nevertheless, mean accuracy was high again (m = 86.04%, $SD = 3.05\%$), with all participants scoring above 78%.

Reaction time (RT) data from Experiments 1 and 2 were also examined. During data collection, it was noted that participants would occasionally pause before responding to a task image, in order to blink or rest their eyes before moving to the subsequent trial. To remove these instances from the examination of RT, all responses greater than 2 standard deviations from each participant's mean RT were dropped. The mean RT in Experiment 1 was 764 ms, similar to that found in the original work by Bemis and Pylkkänen (2011). We also replicated the pattern across conditions observed in the original study, with two-word phrases eliciting the fastest mean RT (m = 722 ms, $SD = 222$ ms), followed by one-word phrases (m = 734 ms, $SD = 218$ ms), one-word lists (m = 744 ms, $SD = 207$ ms), and then

two-word lists (m = 858 ms, $SD = 302$ ms). In Experiment 2, the mean reaction time for two-word phrases was 926 ms. ($SD = 256$ ms). Contents Compounds elicited the longest mean RTs (m = 1,005 ms, $SD = 330$ ms), followed by Material Compounds (m = 986 ms, $SD = 322$ ms), then Color-noun phrases (m = 890 ms, sd = 272 ms). No behavioral data were collected in the letter-string response localizer, as participants passively viewed the stimuli.

3.2 | Letter-string response localizer

Analysis of data collected in the letter-string response localizer successfully replicated the patterns found in by Gwilliams et al. (2016). For each analysis, multiple statistically significant spatiotemporal clusters were found, however we report here only those that matched, most closely, the patterns found in the previous work. The full set of clusters is reported in the Supporting Information. We consider clusters that localized to the same region and showed the same waveform morphology as found by Gwilliams et al. to provide compelling evidence that their specific timing and localization are generalizable. The labels of Noise Level and String Type effects refer to the original characterizations by those authors.

Figure 2 displays the localizer results. For each analysis, cluster-based permutation tests were performed on the time course of regression coefficients. For ease of interpretation, Figure 2 shows the mean response within each relevant condition, computed as the average sensor-level response converted to a noise-normalized source estimate for each participant and subsequently averaged across individuals. In lateral occipital and posterior temporal areas, we observed an initial sensitivity to visual noise properties of the stimuli. A significant effect matching the profile of the original Type I Noise Level effect was found in the lateral and ventral occipital lobe, spanning 100–180 ms after stimulus onset (corrected $p < .0001$) and capturing a negative-going component in response to high noise stimuli that was absent in low noise stimuli. A second Noise Level effect, an opposing dissociation between high and low noise conditions, was found in the posterior fusiform gyrus between 105 and 175 ms (corrected $p < .0001$) with the peak at approximately 130 ms.

More anteriorly, we observed a change in sensitivity to a discrimination between letter and symbol strings. Two clusters in the test window reached the threshold for statistical significance. The first of these was found toward the most anterior point of the ventral ROI and spanned 130–165 ms ($p < .0001$), thus overlapping in time with the more posterior sensitivity to visual noise. The second cluster was located slightly more posterior on the ventral surface (see Figure 2) and later in time, spanning 260–300 ms ($p < .0001$).

Expanding on the analyses performed by Gwilliams et al., and repeated here, we also performed a series of follow-up tests that separated the four- and one-unit letter and symbol-string stimuli. The results are shown in Figure S1. First, in the more anterior and earlier String Type cluster, both individual letter and four-letter words showed a transient increase relative to their symbol-string counterparts, appearing as a larger
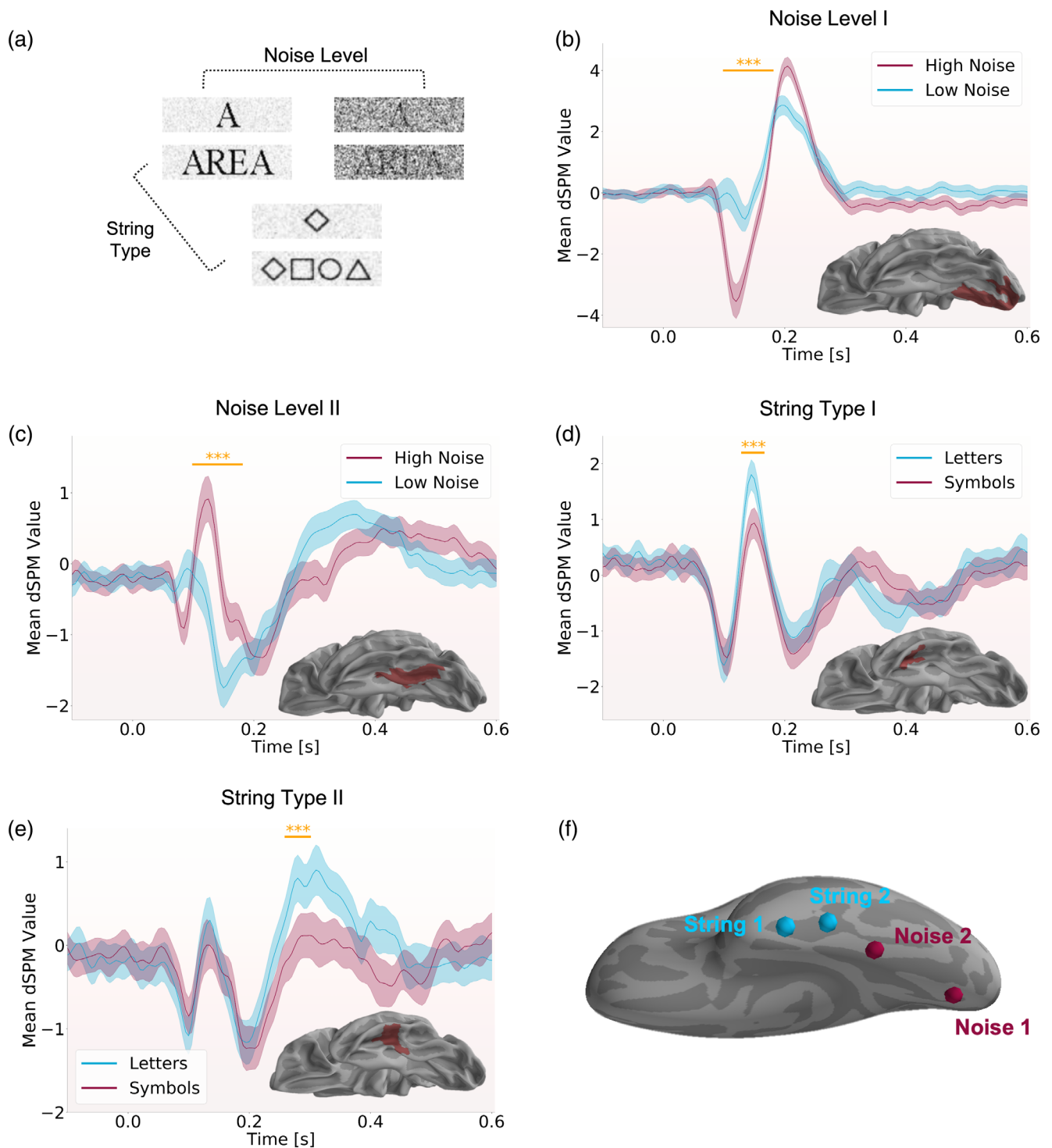
FIGURE 2 Letter-string response localizer. (a) Participants viewed one-letter items embedded in high and low visual noise, four-letter words embedded in high and low visual noise, and one- and four-unit length symbol strings, in low visual noise. (b) A dissociation between high and low visual noise stimuli was observed in left occipital cortex between 100 and 180 ms post stimulus onset ($p <.0001$). (c) In the left posterior fusiform gyrus, high and low noise letter-string stimuli elicited opposite responses between 105 and 175 ms ($p <.0001$). (d) Responses in the left anterior fusiform and inferior temporal gyrus showed an initial dissociation in responses to letter-string stimuli and symbol stimuli (across both one- and four-unit lengths) between 130 and 165 ms after stimulus onset ($p <.0001$). (e) A second dissociation between String Type was found slightly later, between 260 and 300 ms ($p <.0001$) in a cluster that was also found in the anterior fusiform, but located slightly more posterior than the initial String Type dissociation. (f) The center of mass for each cluster is shown on the inflated cortical surface

positive peak in the responses. This relative increase appeared to be larger in the Words versus Symbols contrast than the Letter versus Symbol contrast. The result of a post-hoc paired t-test conducted on the

differences between the contrasts (i.e., Letter—Symbol vs. Word—Symbols) between 100 and 200 ms after stimulus onset was statistically significant at an uncorrected $p <.05$ level ($t[27] = 2.82$, $p <.009$).

A similar pattern was found in the post-hoc examination of the later, but slightly more posterior, String Type cluster. Here, words elicited a more prolonged divergence from their symbol counterparts, relative to that seen between individual letters and symbols. In the Word versus Symbols contrast, this divergence extended beyond the original 300 ms boundary on the original test window. A post-hoc paired *t*-test on the differences between the contrasts, this time between 260 ms (the cluster onset) and 400 ms, again revealed a statistically significant difference ($t[27] = 6.43$, $p < .0001$), with the four-unit contrast eliciting larger differences than the one-unit contrast.

The spatial extents of the clusters that matched the Gwilliams et al. (2016) results were adopted as group-level functional (f)ROIs in the analyses reported below to examine where along this series of responses we might observe a sensitivity to the different word combination conditions. This set of functional ROIs consisted of the Noise Level I and II clusters, and both String Type clusters.

3.3 | Experiment 1: Adjective-noun combinations

Spatiotemporal cluster-based permutation tests were first performed in the three ROIs defined as the left ATL, PTL and AG. Statistically significant composition effects (i.e., a dissociation of the two-word phrase condition from the remaining three) were found in all three ROIs. In the left ATL, significant clusters appeared in the earliest and latest time windows (150–300 ms, $p = .0063$; 450–600 ms, $p = .0135$; all reported *p*-values FDR-corrected) while in the left PTL ROI significant clusters were found in all time windows (150–300 ms, $p = .0006$; 300–450 ms, $p = .0081$; 450–600 ms, $p = .0006$). In the left AG ROI, significant clusters were found in the 300–450 ms ($p = .0081$) and 450–600 ms ($p = .0088$) windows. Figure 3 shows the time-course of activity localized to representative clusters in each ROI. In the ATL and PTL, dissociations between the two-word phrases and the remaining conditions were notably similar and the largest significant cluster in the left
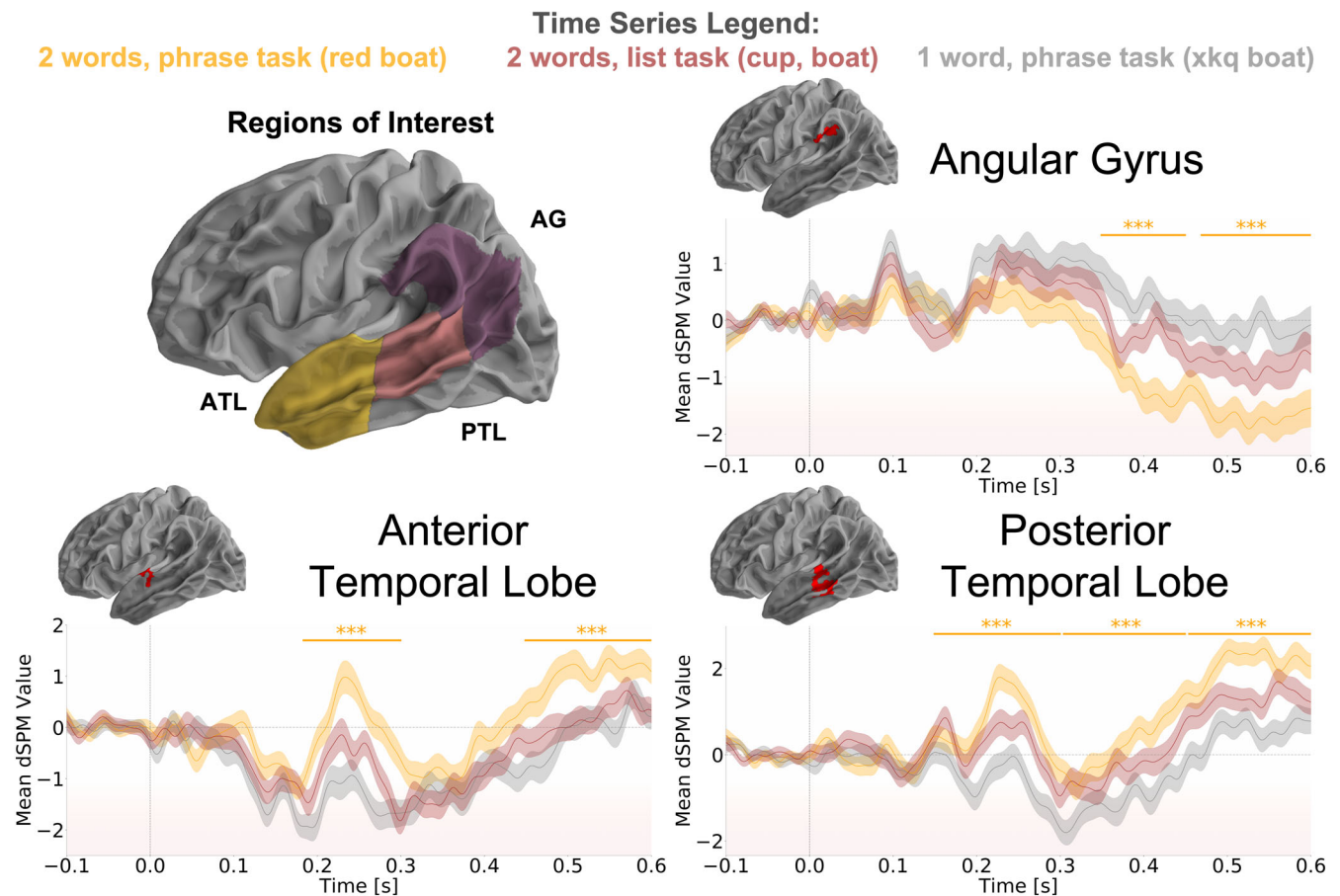

FIGURE 3 Experiment 1. Adjective-noun composition. Responses were examined in three regions of interest (ROIs), top left. Significant composition effects, defined as magnitude increases in response to two-word phrase (e.g., red boat) relative to two-word lists (cup, boat) and one-word baseline materials (xkq boat), were found in all three ROIs. In each time series, 0 ms indicates the onset of the second word on each trial (i.e., the phrasal head). Shown here are representative clusters from each region. Stars and lines indicate significant composition effects. In the ATL (bottom-left), significant effects were found in both the 150–300 ms and 450–600 ms windows, while the AG (top-right) showed the effects in the 300–450 ms, and 450–600 ms windows. The left PTL (bottom-right) showed composition effects in all three analysis time windows. Although the one-word list condition is not shown here for ease of visualization, all composition effects, by test definition, also showed a dissociation from this fourth condition

ATL ROI (shown in Figure 3) was located along the superior temporal gyrus (STG), just adjacent to the left PTL cluster. The spatial extents of these three clusters, one per ROI, were adopted as functional ROIs in the analysis of Experiment 2.

We next examined responses in the functional ROIs defined from the letter-string response localizer, using temporal cluster-based permutation tests to compare responses within each region. Of particular interest was whether the ventral anterior areas that showed a sensitivity to String Type also showed composition effects. Figure 4 displays the results. In the most anterior ROI, which showed the early String Type dissociation at approximately 130 ms, no significant clusters were found in the original analysis windows, which began at 150 ms. However, inspection of the responses in this functional ROI suggested an even earlier dissociation between two-word phrases and the remaining conditions, appearing as a negative peak at approximately 110 ms. A post-hoc cluster test on activity between 50 and 150 ms revealed a significant (though uncorrected) composition effect (cluster: 83–128 ms, $p = .005$).

In the String Type II ROI, which distinguished letters and symbols later in time, significant composition effects were found in all three of the original analysis time windows (150–300 ms, $p = .007$; 300–450 ms, $p = .0002$; 450–600 ms, $p = .0001$). Using cluster-based permutation tests with a two-by-two repeated measures ANOVA to generate the mass univariate statistic, we also confirmed that this area showed a significant effect of the number of words, with two-word stimuli (both lists and phrases) dissociating from their one-word counterparts, in the 450–600 ms analysis window ($p < .0001$). Last, we asked whether the two Noise Level functional ROIs showed a sensitivity to phrasal composition. As highlighted in Figure 4, the Noise Level II area in the posterior fusiform contained a statistically significant dissociation between two-word phrases and the remaining conditions in the late, 450–600 ms time window ($p = .0005$).
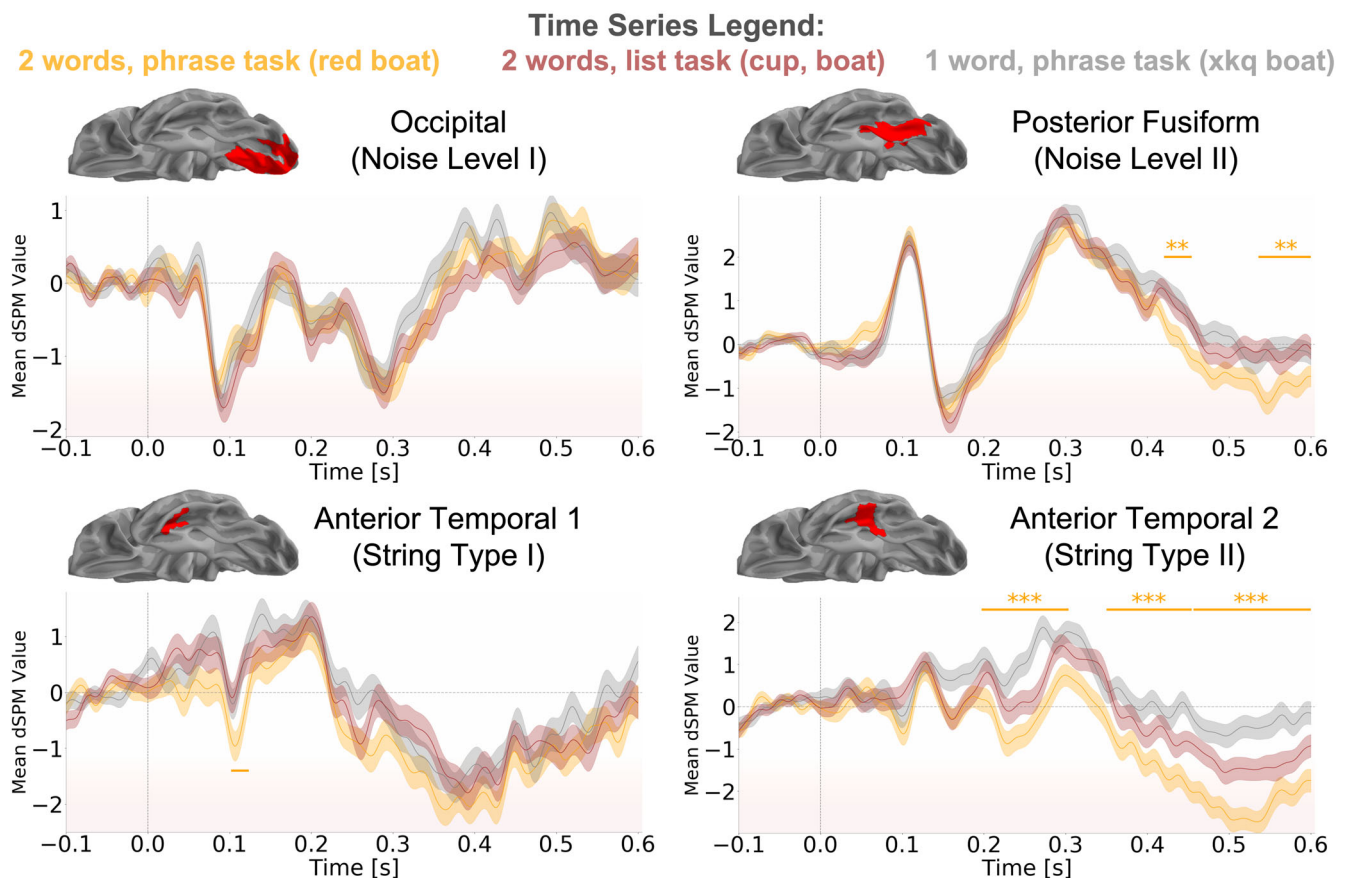
Analyses in letter-string functional ROIs:

FIGURE 4 Experiment 1. Analysis of the functional ROI responses defined from the letter-string response localizer. In the occipital and posterior fusiform ROIs, shown to be sensitive to the level of visual noise in letter- and symbol-string stimuli, all conditions elicited markedly consistent responses. Significant composition effects were found in the posterior fusiform, Noise Level II ROI, but only relatively late in time (300 ms onward). The more anterior, String Type I, ROI displayed a transient dissociation between phrases and the remaining conditions at approximately 100 ms after the onset of the phrasal head, and no subsequent dissociations. This effect was found outside of the a priori analysis windows. In the slightly more posterior, String Type II, ROI, significant composition effects were found in all three analysis windows, spanning 150–600 ms. In each timeseries, 0 ms indicates the onset of the second word on each trial (i.e., the phrasal head). Stars and lines indicate significant effects in the primary analysis windows

In summary, the results of Experiment 1 demonstrated that ROIs capturing the left anterior temporal cortex (150–300 ms, 450–600 ms), posterior temporal cortex (150–600 ms), and AG (300–600 ms) all contained estimated source amplitudes implying a contribution to adjective-noun composition. Additionally, two left anterior fusiform ROIs, defined based on the contrast of letter-strings and symbols in the earlier localizer, also showed increased response magnitudes to two-word phrases relative to the list and single word
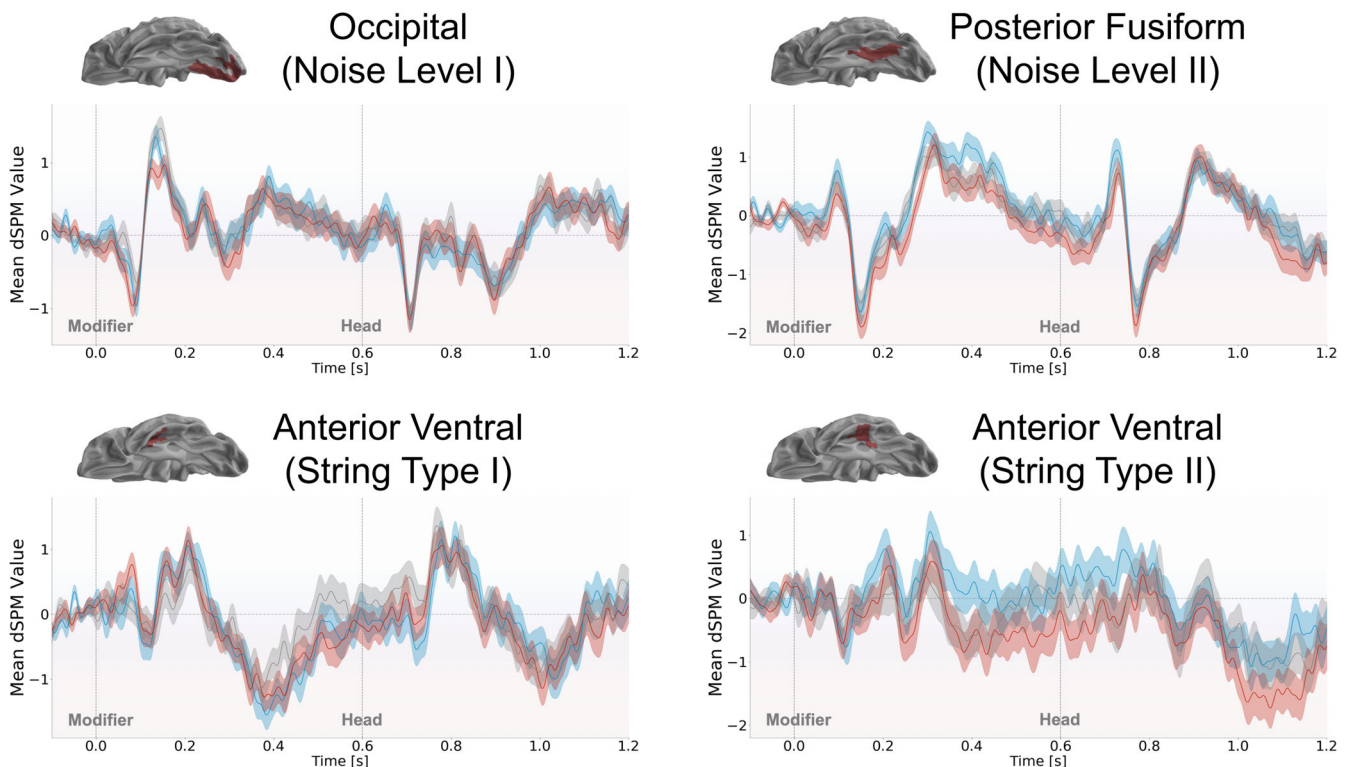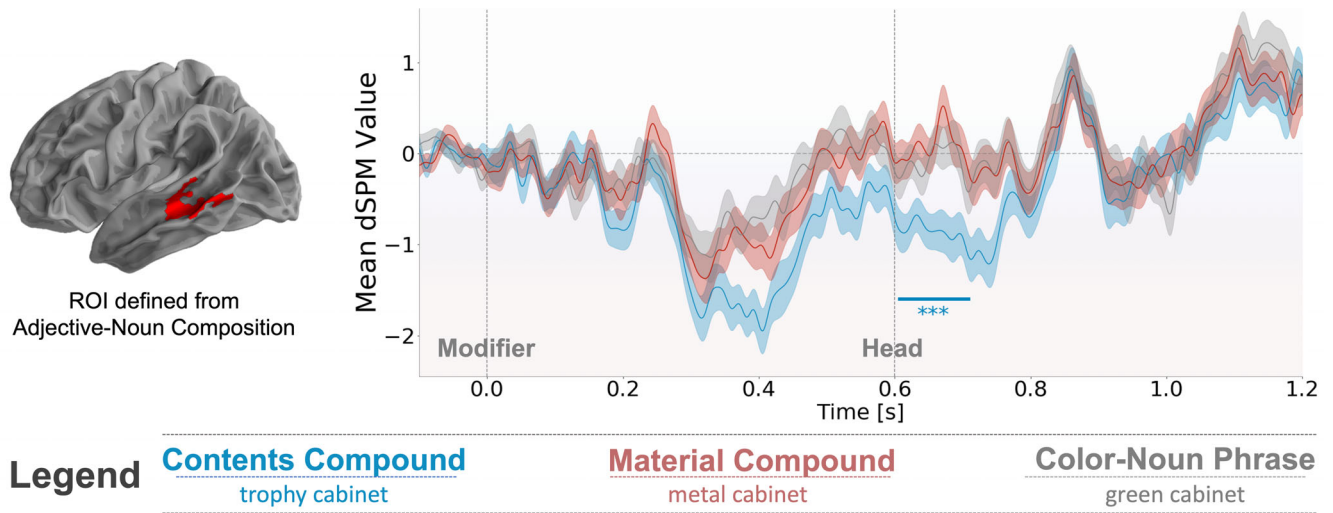

FIGURE 5 Experiment 2. Top: Contents Compounds, involving a spatial or functional relation between the combined nouns, elicited increased response magnitudes in the left posterior temporal ROI relative to both the Material Compounds and Color-Noun Phrases. This pattern appeared to begin in response to the phrasal modifiers and re-emerged for the first 100–200 ms after the onset of the phrasal head. Bottom: No significant differences between the Experiment 2 phrasal stimuli were found in the series of ventral surface ROIs defined from the letter-string response localizer

conditions. The most anterior of these discriminations was transient and appeared only 100 ms after the onset of the phrasal head, while the more posterior discrimination appeared in the same windows as the lateral effects, beginning between 150 and 300 ms after phrasal head and showing a sustained discrimination from then onward. We next used each of these areas as functional ROIs to examine the responses of this phrasal composition network to noun–noun compounds, differing in thematic versus feature-based modifications.

3.4 | Experiment 2: Noun–noun combinations

The analysis of Experiment 2 focused on identifying differences in responses to Contents Compounds, involving a spatial or functional thematic relation between the composing words (e.g., trophy cabinet) and Material Compounds, where the modifier specifies the material the head noun is made of (e.g., metal cabinet). The main results are shown in Figure 5. We began by conducting temporal cluster-based permutation tests in three ROIs defined from the adjective-noun composition clusters that were found in the ATL, PTL, and AG in Experiment 1. This approach used the clusters as functional ROIs, with the intention of increasing sensitivity by targeting a subset of the larger area's sources.

In the left PTL, although no significant clusters were found in the original windows, inspection of the time-course of activity revealed an obvious dissociation between Contents and Material Compounds, appearing before the start of the earliest test period (see Figure 5). We thus conducted an expanded test across the entire 0 to 600 ms window of response to the head noun. This revealed a significant difference between Contents and Material Compounds (cluster: 6 ms to 112 ms after phrasal head, corrected p-value = 0.0378). Post-hoc comparison of the mean activity in this time window also confirmed a statistically significant difference between Contents Compounds and Color-Noun phrases ($t[27] = 3.40$, $p = .002$) and failed to find a difference between Material Compounds and the Color-Noun phrases ($t[27] = 0.69$, $p = .495$).

Examination of the mean time-course in the PTL also suggested an earlier dissociation between Contents Compounds and the other phrasal stimuli, beginning in responses to the modifiers (see Figure 5, top). A follow-up test for differences between Contents and Material Compounds in responses to the phrasal modifiers did reveal a cluster that was statistically significant without accounting for multiple comparisons (368–386 ms relative to modifier onset, $p = .0236$, uncorrected). This finding then raised the question of whether a similar dissociation between Contents and Material modifiers would be found if these words were read in isolation, rather than as phrasal modifiers. To address this, we extracted the PTL region's responses to the set of Contents and Material modifiers, presented in isolation in separate blocks during the Experiment 2 procedure. These responses are shown in Figure S2 and do hint at a similar dissociation as that found in response to the phrasal modifiers in Figure 5. However, a temporal cluster-based test comparing responses to Contents and Material modifiers in

isolation failed to find any clusters that formed between 300 and 600 ms after word onset.

Beyond the left PTL, none of the remaining adjective-noun composition clusters showed significant differences between Contents and Material Compounds. The mean responses in the AG and ATL adjective-noun ROIs are shown in Figure S3. To address the possibility that separate areas of the encompassing regions may show differences, we also performed spatiotemporal tests across each of the three ATL, PTL, and AG regions, as was done in the analysis of Experiment 1. These also failed to find any significant differences between the conditions of Experiment 2. Last, cluster-based tests conducted in the Noise Level and String Type functional ROIs defined from the letter-string response localizer also failed to find any significant differences. The responses of these areas to the full phrases are displayed in the bottom panel of Figure 5, highlighting the uniformity between all phrasal stimuli along the progression of letter-string response areas.

3.5 | Localization accuracy and crosstalk

To examine localization accuracy in the loci of our effects, we first computed the average PLE from the PSFs of sources within each of the spatiotemporal clusters identified in the previous analyses. This was done using the covariance matrices and inverse estimators for all participants from data collected in Experiment 2. The results are shown in Figure 6 (left). Previous examinations of localization error have shown that PLE varies across the cortical surface but tends to be below 5 cm (Hauk et al., 2011; Molins et al., 2008). With some exceptions, the majority of localization errors observed here were below 5 cm, and the mean values across participants fell between 1.28 and 2.00 cm. Along the ventral surface, the general pattern was that more anterior regions had increasing localization errors, consistent with previous results finding particularly high localization uncertainty in the ATL (Hauk et al., 2011; Hillebrand & Barnes, 2002).

In targeted examinations we compared the CTFs for all sources in the three lateral left hemisphere adjective-noun clusters (in the ATL, PTL, and AG), as well as the String Type II cluster that showed adjective-noun composition effects in overlapping temporal windows. Of particular interest was whether responses localized to each of these regions were potentially contaminated by leakage from the others. For examination, the CTFs for constituent sources in each ROI were averaged, converted to absolute values, and masked with a half maximum threshold. These are shown in Figure 6 (right). It should be noted that this choice of threshold was arbitrary. The continuous valued CTFs for each ROI are shown in Figure S4. Inspection of the CTFs demonstrated that the adjective-noun clusters in the left AG and PTL overlapped in their CTFs in and around temporoparietal cortex. The continuous valued CTFs (Figure S4) also revealed this overlap and demonstrated that, to a lesser degree, the left ATL CTF overlapped with that of the String Type II cluster on the ventral surface. These overlaps may account for the similarity in the waveforms seen in the left ATL and String Type II responses in Experiment 1, as well as
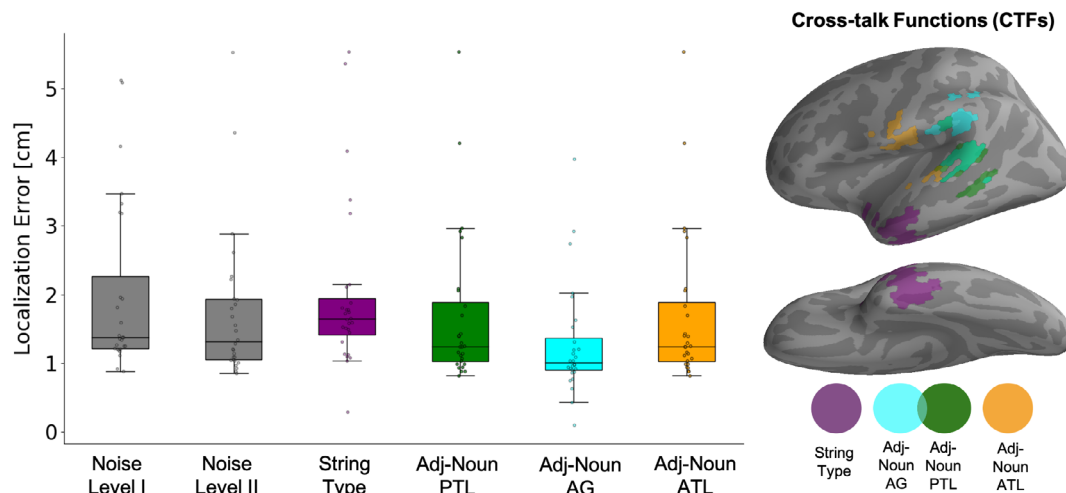
FIGURE 6 Left: Peak localization error (PLE) was calculated as the Euclidean distance between the position of each constituent source and the absolute maximum in that source's PSF. PLE for an ROI was calculated as the average PLE of all constituent sources. Right: Average CTFs for the sources within each region that showed similarly timed adjective-noun composition effects, with half maximum absolute amplitude applied as a mask. The CTFs for the AG and PTL regions showed clear overlap, suggesting that responses localized to each may be conflated. The CTF for the more ventral String Type ROI does not appear to overlap with the more lateral sites when this half amplitude is applied, however, a small degree of overlap with the ATL adjective-noun cluster can be seen in the continuous valued CTFs in Figure S4. String Type indicates the String Type II ROI; Adj-Noun indicates clusters identified in the Experiment 1 contrast of adjective-noun composition; AG, angular gyrus; ATL, anterior temporal lobe; PTL, posterior temporal lobe

the similarity in the late (300 ms onward) response patterns in the left AG and PTL clusters.

We additionally extracted the CTFs for the four letter-string response functional ROIs on the ventral surface. These are shown in Figures S4 (continuous) and S5 (half amplitude threshold). Particularly notable was the extent of the posterior fusiform Noise Level II CTF, which showed the potential influence of activity from areas as far lateral as the posterior superior temporal sulcus, and as far anterior as the String Type functional ROIs. This larger CTF may thus account for the late adjective-noun composition effects that were observed in this ROII (see Figure 4), as they could reflect the smearing of activity patterns generated elsewhere. The half maximum CTF map also shows the overlap among each progressively more anterior CTF, demonstrating that precise localization of the stages of letter-string responses is obfuscated by the limitations of MEG spatial resolution.

3.6 | Whole-surface statistical contrasts

Finally, to ensure that a focus on ROI analyses did not lead to missed dissociations in other cortical regions, we examined statistical contrast maps between the primary conditions of Experiment 1 and 2 across the cortical surfaces. We note that these are primarily descriptive results, not subject to appropriately corrected statistical tests for their robustness. The lateral and ventral surface maps are shown in Figures S6–S8. In the comparison of two-word phrases and two-word lists from Experiment 1, in addition to the superior sections of the left ATL and PTL found in the ROI analyses, dissociations were also seen in the right ATL (200–300 ms), prefrontal cortex (200–300 ms), and

temporal–parietal cortex (500–600 ms). The whole-surface visualizations also make clear the prominence of the ventral ATL dissociation between two-word phrases and two-word lists (as well as the one-word items), which can be seen as a sustained cluster of *t*-values at 200 ms and from 400 ms onward. In the comparison of Contents and Material Compound stimuli of Experiment 2, the left PTL dissociation just after the onset of the head noun was most apparent, but there were also hints of an inferior frontal dissociation between the two conditions from 400 ms onward. The comparison of Contents Compounds and adjective-noun phrases (i.e., trophy cabinet vs. green cabinet) revealed notable differences near the left inferior frontal gyrus (0–200 ms), left and right medial temporal lobe (200–400 ms), and the isthmus of the cingulate gyrus (200–300 ms).

4 | DISCUSSION

This work examined how the comprehension of adjective-noun and noun–noun concepts, which involved feature-based versus thematic modifications, differentially taxed neural responses in putative semantic hub regions, as well as the ventral visual word processing stream. Figure 7 displays a summary of the findings. Consistent with previous studies implicating the left PTL in the processing of thematic knowledge or relationality (e.g., Williams et al., 2017), our results suggest that posterior sections of the left temporal lobe are involved in the recognition and processing of implicit thematic relations between words during visual reading. We also provide evidence suggesting that this same area is engaged in the comprehension of adjective-noun phrases when the adjective modifies only the color of the head noun
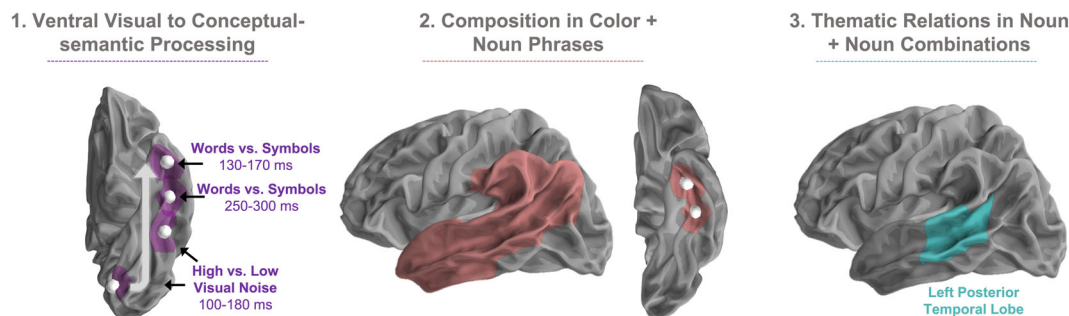
FIGURE 7 Summary of the primary findings. (1) The results of the letter-string response localizer suggest a visual to lexical transformation along the ventral surface of the left hemisphere in support of visual word recognition. White circles indicate the approximate location of each relevant cluster, which showed similar localizations and morphology to those identified by Gwilliams et al. (2016). The grey arrow indicates the inferred flow of information processing, from posterior to anterior. (2) Composition in color + noun phrases elicited the engagement of left perisylvian regions of interest, including the anterior temporal lobe, posterior temporal lobe, and angular gyrus, along with anterior sections of the left fusiform gyrus. Although a large patch of cortex is highlighted, only smaller portions of this may contain the true generators of this effect (see Figures 2–5). (3) Of those areas engaged by color + noun combinations, only the left posterior temporal lobe ROI was further modulated by the need to retrieve or specify thematic relations between composing words in noun–noun combinations

(e.g., red boat), demonstrating that its contribution to composition is not limited to thematic or relational processing. Sections of the left ATL (150–300 ms), anterior fusiform (150–300 ms), and a temporoparietal ROI containing the left AG (300–450 ms), were also engaged by adjective-noun composition.

Altogether, this pattern demonstrates that a distributed left hemisphere network supports composition during reading, even when that composition involves only a simple feature modification (e.g., color + noun). Although prior MEG studies of adjective-noun composition have related parts of this network to minimal composition (most consistently the left ATL; e.g., Bemis & Pylkkänen, 2011; Westerlund & Pylkkänen, 2014; less frequently, the left PTL and/or AG: Bemis & Pylkkänen, 2013), to the best of our knowledge engagement of the larger network as a whole has not been directly reported in any single previous study (see more discussion below). Our results highlight the possibility of this more distributed network, elucidated with MEG, and build on this finding by suggesting that a sub-component, the left PTL, also houses computations related to retrieving or otherwise processing thematic knowledge relevant to the ongoing combination of words. Rather than a strict thematic versus taxonomic distinction between posterior and anterior temporal areas, as predicted by some accounts (e.g., Schwartz et al., 2011), these results may instead indicate the overlap or adjacency of feature and thematic processing in the left PTL, in the context of composition. Our data also demonstrate that sections of the left mid and anterior fusiform house neural activity supporting visual word recognition, which is modulated by the need to combine that incoming word with a preceding modifier. This apparent overlap of lexical and combinatory processing appears most clearly between 150 and 300 ms after word onset. Together with its position at the termination of the visual word processing stream, this suggests that it may be one of the earliest stages of visual word processing that is influenced by combinatory demands. We outline each of these findings in greater detail below, alongside proposed interpretations that we hope will motivate future research.

4.1 | Relational processing modulates responses in the left PTL

Our data provide evidence that a section of left posterior temporal cortex contributes to the combination of color adjectives and simple object-denoting nouns to form phrases (e.g., red boat), as well as the processing of thematic relational structures in noun–noun combinations. This latter characterization stems from the thematic nature of the Contents Compounds (e.g., trophy cabinet) as compared to the Material Compounds (e.g., steel cabinet) and adjective-noun phrases (e.g., green cabinet). Specifically, the spatial or functional relationship between the two nouns in Contents Compounds can be plausibly said to rely on thematic knowledge of the concepts (i.e., co-occurrence or complementary roles of the two objects in an event/schema; Estes, Golonka, & Jones, 2011) and cannot be paraphrased as a simple predication (i.e., a trophy cabinet is not a cabinet that is a trophy). On the other hand, both Material modifiers and the color modifiers in the adjective-noun phrases can be adequately paraphrased as predications and specify simple, predominantly visual, features of the head nouns' denotations. Considered from this perspective, the present result is consistent with some predictions of dual hub accounts of semantic knowledge (de Zubicaray et al., 2013; Schwartz et al., 2011) but not others. The presumed presence of thematic linkages between words only modulated responses in a posterior region of the left temporal lobe, rather than the left anterior temporal regions proposed, in alternative accounts, to serve both thematic and taxonomic knowledge retrieval. However, we also observed that relative to non-combinatory stimuli, feature modifications in adjective-noun phrases (e.g., red + boat) engaged the PTL and the left AG ROIs, demonstrating that the contribution of responses in these areas to compositional operations is not exclusive to thematic information processing. Thus, our results do not straightforwardly support the strictest version of a dual hub account, which would posit that feature-based modifications only modulate left ATL responses.

The localization of the noun–noun compounding effect is also noteworthy for its relation to previous empirical findings. The posterior MTG has been found to house heightened responses to verbs relative to nouns (Bedny et al., 2008; Bedny et al., 2014; Bedny & Thompson-Schill, 2006; Davis et al., 2004; Kable et al., 2002, 2005; Martin et al., 1995; Yu et al., 2011, 2012) while distributed activations in neighboring sections of the left superior temporal sulcus and gyrus have been found to encode relational roles (e.g., woman as agent, girl as patient) in sentence comprehension (Frankland & Greene, 2015; Frankland & Greene, 2020). Assuming verbs activate event and relation knowledge in memory and considering thematic knowledge as being related to how entities co-occur in events or schemata, these findings suggest an underlying thematic or relational processing role for the left PTL. This is consistent with the results of Williams et al. (2017), which offered that relationality, rather than syntactic category or eventivity, is the underlying factor that drives these effects. Nevertheless, in their comparison of relational (i.e., thematic) and attributive noun–noun compounds, Boylan et al. (2017) found greater hemodynamic response magnitudes for relational compounds in the left and right angular gyri, building on previous work that showed distributed activation patterns in the AG correlated with similarity ratings of verb phrases (Boylan, Trueswell, & Thompson-Schill, 2015).

On first consideration, our localization of a Compound Type effect to the posterior MTG appears inconsistent with the results of Boylan et al., leading to questions about differences in the designs, especially the use of novel versus familiar noun–noun compounds, and the sensitivity of fMRI versus MEG. However, this issue is complicated by the overlapping CTFs for responses in the left PTL and AG ROIs, which suggest that the PTL localized response could be generated in either, or shared across both, of the two regions. One way that future work will inform our understanding of this difference is by accounting for design differences and directly comparing familiar and unfamiliar noun–noun combinations, using a common neuroimaging modality. The present finding that the left PTL shows dissociations between thematic and nonthematic noun–noun compounds, on both the modifier and head noun, suggests that the added temporal resolution of electrophysiological methods, like MEG, may be important in future study. A recent MEG examination of adjective-noun composition demonstrated that representations of adjectives appear to be active during the processing of a subsequent phrasal head noun (Fyshe, Sudre, Wehbe, Rafidi, & Mitchell, 2019), possibly supporting their combination. Psycholinguistic accounts of English noun-noun combinations and relational structures have highlighted the potentially privileged role of the modifier in prespecifying a thematic relation between the two constituents, which is then evaluated for appropriateness when processing the subsequent head noun (Gagné, 2002). This propose and confirm process is thus one potential account of the left PTL activity observed here, on each constituent noun.

Lastly, the finding that both feature-based color + noun and thematic noun + noun combinations engaged the left PTL, relative to their corresponding baseline conditions, raises the question of whether there are shared or distinct computations underlying the region's involvement in each case. On the one hand, as summarized above, there is now a large body of converging evidence behind the hypothesis that sections of the PTL play a role in relational and/or thematic knowledge processing. However, there is also evidence for the left PTL's involvement in syntactic processing in language comprehension (Flick & Pylkkänen, 2020; Matchin, Brodbeck, Hammerly, & Lau, 2019; Matchin, Hammerly, & Lau, 2017; Rodd, Longe, Randall, & Tyler, 2010; Rogalsky et al., 2018; Snijders et al., 2009; Tyler, Cheung, Devereux, & Clarke, 2013). Syntactic processing could account for the PTL's differentiation of two-word phrases versus lists and single words, since the latter two conditions do not require the construction of phrasal structure. This said, the absence of such effects in some previous studies of phrasal composition (e.g., Bemis & Pylkkänen, 2011), would remain a puzzle (see more discussion below). It is also intriguing to consider that thematic or relational noun–noun compounds could place greater demands on lexico-syntactic processing relative to nonrelational noun–noun combinations, and this could account for part of their increased activation. This may be the case if, for example, accessing the covert relations between nouns requires specification of an argument structure related to knowledge of an intervening verb (e.g., the cabinet *containing* the trophies).

4.2 | Combinatory context influences visual letter string responses

Although the primary focus of this investigation was the processing of semantic relations between words, our results also suggest implications for how the visual processing of words and letter-strings feeds into composition operations. Consistent with original findings from the development of the letter-string response localizer (Gwilliams et al., 2016 following Tarkiainen et al., 1999; and see Neophytou et al., 2018), we identified a series of responses along the ventral surface of the left hemisphere that appear to underpin visual letter and word processing. These initially discriminated between high and low visual noise properties of the stimuli, in the occipital lobe and posterior fusiform, and then differentiated letter- and symbol-string stimuli in the middle to anterior fusiform gyrus, with words diverging from length-matched symbol strings to a greater degree than individual letters. This sequence of responses thus replicates previous evidence (e.g., Gwilliams et al., 2016; Vinckier et al., 2007; Woolnough et al., 2020) for a transition along the ventral surface of left occipital and temporal cortex from visual and orthographic to lexical processing.

At the more anterior sites of this sequence, in the mid and anterior fusiform, the timings of the effects were noteworthy for two reasons. First, while Gwilliams et al. (2016) observed a single String Type effect between 150 and 200 ms (which also showed a later component between 200–300 ms), our data contained two loci showing String Type dissociations, with the earlier (String Type I) and more anterior cluster appearing at approximately 130–160 ms, followed by a second (String Type II) slightly posterior cluster at approximately

260 ms (though we caution over-interpretation of these exact timings and localizations, which may not generalize). These two loci also differed in their sensitivities to the adjective-noun stimuli of Experiment 1. The first demonstrated a relatively early (approximately 100 ms after word onset) and transient discrimination of phrases, lists, and words (see Figure 4), similar to its dissociation of letters and symbols, while the more posterior cluster showed adjective-noun composition effects in all three of the original time windows, spanning 150–600 ms after the phrasal head. One potential account of the earlier String Type I effect is that it reflects a rapid feedforward response from primary visual cortex, sensitive to both lexical and combinatory demands, which may initialize a complementary anterior-to-posterior flow of processing in the ventral visual word pathway (see Woolnough et al., 2020 for a similar proposal regarding lexical processing). This remains speculation for the time being, and future work is needed to confirm whether this spatiotemporal pattern can be replicated.

Second, the duration of the anterior fusiform's discrimination between letters and symbol strings, in the String Type II ROI, raises questions about what type of computations would result in divergences this long. Figure S1 demonstrates that these latencies differed between the letter and word contrasts with their length-matched symbol strings, such that words diverged from symbols longer than the letter stimuli. In an examination of mid-fusiform responses to words and word-like stimuli, Woolnough et al. (2020) found, in sentence reading, responses to high frequency words diverged from pseudowords by approximately 180 ms after onset, for a duration of upward of 300 ms. Low frequency words, on the other hand, did not diverge from pseudowords until later in the epoch, which the authors interpreted as evidence that the mid-fusiform maps visually presented word inputs to entries in the mental lexicon. Our findings are consistent with this proposal, and further implicate the site of this mapping, the left mid and/or anterior fusiform gyrus, in the combination of word meanings. We thus highlight this as a tentative site where lexical access may interact with combinatory demands, enabling the meaning of a word, as it is visually recognized, to quickly feed into composition operations.

The finding that responses in this fusiform region did not differentiate the two noun–noun compound types (Material vs. Content Compounds) offers that it may represent a relatively early stage in combinatory processing, able to distinguish only between some compositional and noncompositional stimuli, but not subtleties among them. This is consistent with the results of Ziegler and Pylkkänen (2016), which found that the left ATL's contribution to phrasal composition, 150–200 ms after the phrasal head, was limited to instances where the modifier did not require in-depth processing of the head noun (i.e., intersective adjectives such as *dead* which have context-independent meanings). Relatedly, previous electrophysiological findings have demonstrated that, slightly earlier than the timing of our String Type II effect, a feedforward pass of activity along the ventral surface of the left hemisphere supports a transformation from visual word input to coarse semantic category information (Chan et al., 2011). This may be reflected in the earlier and transient String Type I response observed at approximately 130 ms.

4.3 | Distributed composition responses to minimal phrases

Our data provide new evidence for overlap between visual word recognition and combinatory processing in ventral sections of the left ATL. However, this area was not the only location at which we observed heightened responses to compositional color-noun phrases relative to noncompositional word lists and one-word items. The same response pattern was also found in sections of the left lateral ATL, PTL, and the ROI containing the left AG, suggesting a distributed compositional network. Previous MEG studies of minimal composition have not always found this wider spread of effects across the lateral left hemisphere, even though a similar network has been implicated by hemodynamic studies examining larger instances of compositional language (i.e., sentences or parts thereof; Fedorenko, Hsieh, Nieto-Castañón, Whitfield-Gabrieli, & Kanwisher, 2010; Matchin et al., 2017; Pallier, Devauchelle, & Dehaene, 2011). Instead, the reported minimal composition MEG effects have been largely constrained to the left ATL. For example, Bemis and Pylkkänen (2011) tested the left ATL and a single ROI encompassing both the pMTG and AG, finding that only the former showed a composition effect. Kim and Pylkkänen (2019), examining minimal verb phrases, also failed to find any effects in temporoparietal regions.

There have, however, been hints of this wider left hemisphere network, even in minimal composition MEG designs. Bemis and Pylkkänen (2013) found that the left AG showed heightened responses to composition in both reading and auditory comprehension. Later studies, which examined modifier-noun phrases but varied the nature of the stimuli (e.g., Westerlund & Pylkkänen, 2014; Zhang & Pylkkänen, 2015), did not explicitly examine the responses localized to posterior regions of the temporal lobe or AG. Here, however, closer examination suggests that at least some of these studies may have captured a distributed composition response. For instance, the spatiotemporal results of Flick et al. (2018) suggested that a larger portion of the left temporal lobe, including both anterior and posterior superior temporal, and ventral anterior areas, were engaged by phrasal composition. A similar pattern was observed in the surface contrast maps of phrases versus single words by Westerlund and Pylkkänen (2014). We therefore conservatively believe that the issues of whether the left PTL and AG are engaged in minimal composition, and whether this can be observed with MEG, are not yet settled. The current results add support to the notion that both areas, together with the ATL, are involved in this process, implicating a larger section of perisylvian cortex in minimal combinatory processing. The present study had many potentially important methodological differences from these earlier works, including the use of anatomical MRIs to construct forward models, and the orientation constraint placed on source estimates. A complete analysis of the impact of each of these factors is beyond the scope of this paper but represents an important topic for future work.

4.4 | The impact of signal leakage and other limitations

The use of source estimation in MEG data analysis comes with important caveats. In our case, examination of the spatial uncertainty in our

estimators demonstrated the expected blurring or signal leakage, which was particularly notable across left lateral temporal lobe areas. This has consequences for the interpretation of responses localized in these areas, discussed in previous sections. Clusters found in the lateral ROIs containing the AG and PTL showed clear overlap in their CTFs (Figure 6, right). For each constituent source within a parcel or cluster, CTFs reflect the possible influence of all other sources on amplitude estimated at that source. Thus, overlapping CTFs highlight that, rather than independent responses, activity localized to each area may reflect a smearing in the transformation of MEG responses from sensor to source space. Because of this possible smearing, we cannot make strong claims about the precise localization of the left PTL and AG contributions to adjective-noun composition, which may arise in either, or both, of the two areas. However, the previous fMRI findings of distributed activation in language comprehension, across a network that includes the left ATL, PTL, and AG (Fedorenko et al., 2010; Matchin et al., 2017; Pallier et al., 2011), conflict with the idea that the distributed source patterns observed here are entirely attributable to signal leakage.

The inspection of continuous valued CTFs (Figure S4) also revealed overlap between the loci of adjective-noun composition effects in the anterior fusiform (i.e., the String Type ROIs) and the lateral ATL cluster found in Experiment 1. This raises questions about whether each of these responses is an independent contribution to adjective-noun composition, and about the true localization of the previously observed left ATL composition responses. It is possible, for example, that the ventral anterior temporal region that was identified by the String Type contrast of the letter-string localizer is the true generator of the left ATL composition response, which shows up in more lateral sites due to signal leakage. As mentioned above, a subset of previous MEG studies have observed a more distributed spatiotemporal effect, including both superior and ventral sections of the left ATL (e.g., Flick et al., 2018; and see the whole-cortex contrast maps of Bemis & Pylkkänen, 2011; Westerlund & Pylkkänen, 2014). The present results highlight that this could be another artifact of overlapping CTFs or an indication of separate contributions to composition. Their dependence on data means that CTFs should be expected to fluctuate from dataset to dataset, and the degree of variability across studies has not yet been characterized. This variability, we speculate, may also account for some of the inconsistencies across previous MEG studies of phrasal composition, which have identified relevant effects in different sections of the ATL (e.g., Bemis & Pylkkänen, 2011; Flick et al., 2018; Kim & Pylkkänen, 2019; Westerlund & Pylkkänen, 2014; Zhang & Pylkkänen, 2015). Critically, none of this insight would be available without the examination of crosstalk and point-spread functions, reinforcing the notion that these provide a valuable source of information about the quality of MEG source estimators, and should be routinely examined (Hauk et al., 2011, 2019). These findings also highlight the caution that should be applied not only in the interpretation of the present results, but also those of future MEG studies that use source estimation with the goal of identifying functional subregions, particularly in the ATLs where localization accuracy may be lower.

Another limitation of the present study is related to the number of semantic relations between composing words that were examined in this stimulus set, limited to only "head made of modifier" and "head containing modifier" relations, as well as color-noun composition. Previous psycholinguistic research (e.g., Gagné & Shoben, 1997; more recently, Schmidtke, Gagné, Kuperman, & Spalding, 2018), has examined a much larger number of these semantic relations and pointed out important nuances in them, such as the construal of a constituent concept (e.g., plastic squirrel involves a construal of a squirrel, rather than a live one; Wisniewski, 1996). It will be important for future work to investigate how various relations, requiring different aspects of conceptual knowledge of the constituents, tax neural responses in the left PTL and/or if these recruit additional cortical areas.

Finally, we note that the present results should not be taken to suggest that only the relatively small number of ROIs examined here underpin all the computations that enable successful comprehension of multiword concepts or contribute, exclusively, to the precise computations examined (i.e., retrieval of relevant thematic knowledge). Indeed, the uncorrected whole-surface contrast maps (Figures S6–S8) suggest the contribution of other cortical areas in each type of composition. Adjective-noun phrases also appeared to dissociate from word lists within the right ATL and prefrontal cortex, two findings that have been occasionally reported in previous studies (Bemis & Pylkkänen, 2011, 2013), as well as sections of the right temporal–parietal junction. Contents Compounds appeared to elicit relatively greater estimated response magnitudes in medial sections of the temporal lobes (not captured in the letter-string response localizer ROI), and also dissociated from adjective-noun phrases based on responses localized to the isthmus of the cingulate gyrus. Both regions have been previously implicated in the processing of noun–noun combinations (Graves, Binder, Desai, Conant, & Seidenberg, 2010), and we leave further refinement of their functional characterizations to future work.

5 | CONCLUSIONS AND FUTURE DIRECTIONS

Understanding a combination of words often requires inserting unstated semantic material between constituents to specify the complete meaning. The present findings suggest that, in reading, when the interpretation of familiar noun–noun combinations involves an implicit thematic relation (e.g., a horse barn is a barn where horses are kept), retrieving or otherwise processing this relation leads to the engagement of left PTL areas. This left posterior temporal region also appears to be involved in composing the meaning of more straightforward adjective-noun combinations, involving only modification of a concrete noun's color (e.g., red boat). While this minimal composition response has been most consistently observed in anterior sections of the left temporal lobe (e.g., Bemis & Pylkkänen, 2011; c.f., Flick et al., 2018), the present data suggest that a more distributed left hemisphere network, including superior ATL and PTL sites, anterior portions of the fusiform gyrus, and cortex around the left AG, may

support minimal instances of composition. Understanding why this more distributed composition pattern is observed in some instances, but not others, is thus an important avenue for future work.

Our results additionally position the engagement of the lateral combinatorial areas with respect to visual word processing in occipital and ventral temporal sites of the left hemisphere. Previous and present evidence suggests that, in visual reading, combinatorial operations are first supported by a transformation from orthographic to conceptual-semantic representations that takes place in left lateral occipital, and posterior and anterior fusiform areas. Our data demonstrate that portions of this ventral processing stream show remarkably consistent responses across sequential MEG experiments. They also suggest that anterior fusiform or inferior temporal areas involved in early lexical processing may additionally contribute to or be modulated by the process of combining word meanings. We hope that future work will further test and refine these proposals by examining word combinations that exemplify a greater variety of semantic relations and manipulating other theoretically relevant properties in the context of visual word recognition and conceptual combination.

CONFLICT OF INTEREST

The authors declare no conflict of interest in the completion of this work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Graham Flick https://orcid.org/0000-0001-9183-4040

REFERENCES

Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., & Ochiai, M. (2001). Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. *IEEE Transactions on Applied Superconductivity*, *11*(1), 669–672. https://doi.org/10.1109/77.919433

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Baron, S. G., Thompson-Schill, S. L., Weber, M., & Osherson, D. (2010). An early stage of conceptual combination: Superimposition of constituent concepts in left anterolateral temporal lobe. *Cognitive Neuroscience*, *1*(1), 44–51. https://doi.org/10.1080/17588920903548751

Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., & Saxe, R. (2008). Concepts are more than percepts: the case of action verbs. *Journal of Neuroscience*, *28*(44), 11347–11353. https://doi.org/10.1523/JNEUROSCI.3039-08.2008

Bedny, M., Dravida, S., & Saxe, R. (2014). Shindigs, brunches, and rodeos: The neural basis of event words. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(3), 891–901. https://doi.org/10.3758/s13415-013-0217-z

Bedny, M., & Thompson-Schill, S. L. (2006). Neuroanatomically separable effects of imageability and grammatical class during single-word comprehension. *Brain and Language*, *98*(2), 127–139. https://doi.org/10.1016/j.bandl.2006.04.008

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

Bemis, D. K., & Pylkkänen, L. (2013). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, *23*(8), 1859–1873. https://doi.org/10.1093/cercor/bhs170

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Boylan, C., Trueswell, J. C., & Thompson-Schill, S. L. (2015). Compositionality and the angular gyrus: A multi-voxel similarity analysis of the semantic composition of nouns and verbs. *Neuropsychologia*, *78*, 130–141. https://doi.org/10.1016/j.neuropsychologia.2015.10.007

Boylan, C., Trueswell, J. C., & Thompson-Schill, S. L. (2017). Relational vs. attributive interpretation of nominal compounds differentially engages angular gyrus and anterior temporal lobe. *Brain and Language*, *169*, 8–21. https://doi.org/10.1016/j.bandl.2017.01.008

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

Chen, Y., Shimotake, A., Matsumoto, R., Kunieda, T., Kikuchi, T., Miyamoto, S., ... Ralph, M. L. (2016). The 'when' and 'where' of semantic coding in the anterior temporal lobe: Temporal representational similarity analysis of electrocorticogram data. *Cortex*, *79*, 1–13. https://doi.org/10.1016/j.cortex.2016.02.015

Clarke, A., Taylor, K. I., & Tyler, L. K. (2011). The evolution of meaning: spatio-temporal dynamics of visual object recognition. *Journal of Cognitive Neuroscience*, *23*(8), 1887–1899. https://doi.org/10.1162/jocn.2010.21544

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, *125*(5), 1054–1069. https://doi.org/10.1093/brain/awf094

Coutanche, M. N., & Thompson-Schill, S. L. (2015). Creating concepts from converging features in human cortex. *Cerebral Cortex*, *25*(9), 2584–2593. https://doi.org/10.1093/cercor/bhu057

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

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, *26*(1), 55–67. https://doi.org/10.1016/S0896-6273(00)81138-1

Dale, A. M., & Sereno, M. I. (1993). Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. *Journal of Cognitive Neuroscience*, *5*(2), 162–176. https://doi.org/10.1162/jocn.1993.5.2.162

Davey, J., Cornelissen, P. L., Thompson, H. E., Sonkusare, S., Hallam, G., Smallwood, J., & Jefferies, E. (2015). Automatic and controlled semantic retrieval: TMS reveals distinct contributions of posterior middle temporal gyrus and angular gyrus. *Journal of Neuroscience*, 35(46), 15230–15239. https://doi.org/10.1523/JNEUROSCI.4705-14.2015

Davey, J., Thompson, H. E., Hallam, G., Karapanagiotidis, T., Murphy, C., De Caso, I., ... Jefferies, E. (2016). Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. *NeuroImage*, 137, 165–177. https://doi.org/10.1016/j.neuroimage.2016.05.051

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14(2), 159–190. https://doi.org/10.1075/ijcl.14.2.02dav

Davis, M. H., Meunier, F., & Marslen-Wilson, W. D. (2004). Neural responses to morphological, syntactic, and semantic properties of single words: an fMRI study. *Brain and Language*, 89(3), 439–449. https://doi.org/10.1016/S0093-934X(03)00471-1

de Zubicaray, G. I., Hansen, S., & McMahon, K. L. (2013). Differential processing of thematic and categorical conceptual relations in spoken word production. *Journal of Experimental Psychology: General*, 142(1), 131. https://doi.org/10.1037/a0028717

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., ... Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, 330(6009), 1359–1364. https://doi.org/10.1126/science.1194140

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

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

Estes, Z. (2003). A tale of two similarities: Comparison and integration in conceptual combination. *Cognitive Science*, 27(6), 911–921. https://doi.org/10.1207/s15516709cog2706_4

Estes, Z., Golonka, S., & Jones, L. L. (2011). Thematic thinking: The apprehension and consequences of thematic relations. In *Psychology of learning and motivation* (Vol. 54, pp. 249–294). Cambridge, MA: Academic Press. https://doi.org/10.1016/B978-0-12-385527-5.00008-5

Fedorenko, E., Hsieh, P. J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, 104(2), 1177–1194. https://doi.org/10.1152/jn.00032.2010

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9(2), 195–207. https://doi.org/10.1006/nimg.1998.0396

Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., ... Dale, A. M. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14(1), 11–22. https://doi.org/10.1093/cercor/bhg087

Flick, G., Oseki, Y., Kaczmarek, A. R., Al Kaabi, M., Marantz, A., & Pylkkänen, L. (2018). Building words and phrases in the left temporal lobe. *Cortex*, 106, 213–236. https://doi.org/10.1016/j.cortex.2018.06.004

Flick, G., & Pylkkänen, L. (2020). Isolating syntax in natural language: MEG evidence for an early contribution of left posterior temporal cortex. *Cortex*, 127, 42–57. https://doi.org/10.1016/j.cortex.2020.01.025

Frankland, S. M., & Greene, J. D. (2015). An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proceedings of the National Academy of Sciences*, 112(37), 11732–11737. https://doi.org/10.1073/pnas.1421236112

Frankland, S. M., & Greene, J. D. (2020). Two ways to build a thought: distinct forms of compositional semantic representation across brain regions. *Cerebral Cortex*, 30(6), 3838–3855. https://doi.org/10.1093/cercor/bhaa001

Fyshe, A., Sudre, G., Wehbe, L., Rafidi, N., & Mitchell, T. M. (2019). The lexical semantics of adjective–noun phrases in the human brain. *Human Brain Mapping*, 40(15), 4457–4469. https://doi.org/10.1002/hbm.24714

Gagné, C. L. (2002). Lexical and relational influences on the processing of novel compounds. *Brain and Language*, 81(1–3), 723–735. https://doi.org/10.1006/brln.2001.2559

Gagné, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(1), 71. https://doi.org/10.1037/0278-7393.23.1.71

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., ... Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7, 267. https://doi.org/10.3389/fnins.2013.00267

Graves, W. W., Binder, J. R., Desai, R. H., Conant, L. L., & Seidenberg, M. S. (2010). Neural correlates of implicit and explicit combinatorial semantic processing. *NeuroImage*, 53(2), 638–646. https://doi.org/10.1016/j.neuroimage.2010.06.055

Gwilliams, L., Lewis, G. A., & Marantz, A. (2016). Functional characterisation of letter-specific responses in time, space and current polarity using magnetoencephalography. *NeuroImage*, 132, 320–333. https://doi.org/10.1016/j.neuroimage.2016.02.057

Hallam, G. P., Thompson, H. E., Hymers, M., Millman, R. E., Rodd, J. M., Ralph, M. A. L., ... Jefferies, E. (2018). Task-based and resting-state fMRI reveal compensatory network changes following damage to left inferior frontal gyrus. *Cortex*, 99, 150–165. https://doi.org/10.1016/j.cortex.2017.10.004

Hallam, G. P., Whitney, C., Hymers, M., Gouws, A. D., & Jefferies, E. (2016). Charting the effects of TMS with fMRI: Modulation of cortical recruitment within the distributed network supporting semantic control. *Neuropsychologia*, 93, 40–52. https://doi.org/10.1016/j.neuropsychologia.2016.09.012

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2), 413. https://doi.org/10.1103/RevModPhys.65.413

Hauk, O., Stenroos, M., & Treder, M. (2019). EEG/MEG source estimation and spatial filtering: the linear toolkit. In S. Supek & C. Aine (Eds.), *Magnetoencephalography*. Cham: Springer. https://doi.org/10.1007/978-3-030-00087-5_85

Hauk, O., Wakeman, D. G., & Henson, R. (2011). Comparison of noise-normalized minimum norm estimates for MEG analysis using multiple resolution metrics. *NeuroImage*, 54(3), 1966–1974. https://doi.org/10.1016/j.neuroimage.2010.09.053

Hillebrand, A., & Barnes, G. R. (2002). A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. *NeuroImage*, 16(3), 638–650. https://doi.org/10.1006/nimg.2002.1102

Humphreys, G. F., & Lambon Ralph, M. A. (2015). Fusion and fission of cognitive functions in the human parietal cortex. *Cerebral Cortex*, 25(10), 3547–3560. https://doi.org/10.1093/cercor/bhu198

Jackson, R. L. (2020). The neural correlates of semantic control revisited. *NeuroImage*, 224, 117444. https://doi.org/10.1016/j.neuroimage.2020.117444

Jefferies, E., Thompson, H., Cornelissen, P., & Smallwood, J. (2020). The neurocognitive basis of knowledge about object identity and events: dissociations reflect opposing effects of semantic coherence and control. *Philosophical Transactions of the Royal Society B*, *375*(1791), 20190300. https://doi.org/10.1098/rstb.2019.0300

Kable, J. W., Kan, I. P., Wilson, A., Thompson-Schill, S. L., & Chatterjee, A. (2005). Conceptual representations of action in the lateral temporal cortex. *Journal of Cognitive Neuroscience*, *17*(12), 1855–1870. https://doi.org/10.1162/089892905775008625

Kable, J. W., Lease-Spellmeyer, J., & Chatterjee, A. (2002). Neural substrates of action event knowledge. *Journal of Cognitive Neuroscience*, *14*(5), 795–805. https://doi.org/10.1162/08989290260138681

Kalénine, S., & Buxbaum, L. J. (2016). Thematic knowledge, artifact concepts, and the left posterior temporal lobe: Where action and object semantics converge. *Cortex*, *82*, 164–178. https://doi.org/10.1016/j.cortex.2016.06.008

Kalénine, S., Mirman, D., Middleton, E. L., & Buxbaum, L. J. (2012). Temporal dynamics of activation of thematic and functional knowledge during conceptual processing of manipulable artifacts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1274. https://doi.org/10.1037/a0027626

Kalénine, S., Peyrin, C., Pichat, C., Segebarth, C., Bonthoux, F., & Baciu, M. (2009). The sensory-motor specificity of taxonomic and thematic conceptual relations: A behavioral and fMRI study. *NeuroImage*, *44*(3), 1152–1162.

Kemmerer, D., Rudrauf, D., Manzel, K., & Tranel, D. (2012). Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex*, *48*(7), 826–848. https://doi.org/10.1016/j.cortex.2010.11.001

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

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. https://doi.org/10.1038/nrn.2016.150

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4

Lin, F. H., Belliveau, J. W., Dale, A. M., & Hämäläinen, M. S. (2006). Distributed current estimates using cortical orientation constraints. *Human Brain Mapping*, *27*(1), 1–13. https://doi.org/10.1002/hbm.20155

Liu, A. K., Dale, A. M., & Belliveau, J. W. (2002). Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping*, *16*(1), 47–62. https://doi.org/10.1002/hbm.10024

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., ... Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences*, *113*(44), 12574–12579. https://doi.org/10.1073/pnas.1608282113

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

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, *270*(5233), 102–105. https://doi.org/10.1126/science.270.5233.102

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. https://doi.org/10.1002/hbm.24403

Matchin, W., Hammerly, C., & Lau, E. (2017). The role of the IFG and pSTS in syntactic prediction: Evidence from a parametric study of hierarchical structure in fMRI. *Cortex*, *88*, 106–123. https://doi.org/10.1016/j.cortex.2016.12.010

Mirman, D., & Graziano, K. M. (2012a). Individual differences in the strength of taxonomic versus thematic relations. *Journal of Experimental Psychology: General*, *141*(4), 601. https://doi.org/10.1037/a0026451

Mirman, D., & Graziano, K. M. (2012b). Damage to temporo-parietal cortex decreases incidental activation of thematic relations during spoken word comprehension. *Neuropsychologia*, *50*(8), 1990–1997. https://doi.org/10.1016/j.neuropsychologia.2012.04.024

Mirman, D., Landrigan, J. F., & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*, *143*(5), 499. https://doi.org/10.1037/bul0000092

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

Moss, H. E., Rodd, J. M., Stamatakis, E. A., Bright, P., & Tyler, L. K. (2005). Anteromedial temporal cortex supports fine-grained differentiation among objects. *Cerebral Cortex*, *15*(5), 616–627. https://doi.org/10.1093/cercor/bhh163

Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language*, *29*(3), 259–288. https://doi.org/10.1016/0749-596X(90)90001-G

Neophytou, K., Manouilidou, C., Stockall, L., & Marantz, A. (2018). Syntactic and semantic restrictions on morphological recomposition: MEG evidence from Greek. *Brain and Language*, *183*, 11–20. https://doi.org/10.1016/j.bandl.2018.05.003

Noonan, K. A., Jefferies, E., Visser, M., & Lambon Ralph, M. A. (2013). Going beyond inferior prefrontal involvement in semantic control: Evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *Journal of Cognitive Neuroscience*, *25*(11), 1824–1850. https://doi.org/10.1162/jocn_a_00442

Okada, Y. C., Wu, J., & Kyuhou, S. (1997). Genesis of MEG signals in a mammalian CNS structure. *Electroencephalography and Clinical Neurophysiology*, *103*(4), 474–485. https://doi.org/10.1016/S0013-4694(97)00043-6

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. https://doi.org/10.1073/pnas.1018711108

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

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*, *375*(1791), 20190299. https://doi.org/10.1098/rstb.2019.0299

Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language*, *81*(1–3), 666–678. https://doi.org/10.1006/brln.2001.2555

Rodd, J. M., Longe, O. A., Randall, B., & Tyler, L. K. (2010). The functional organisation of the fronto-temporal language system: evidence from syntactic and semantic ambiguity. *Neuropsychologia*, *48*(5), 1324–1335. https://doi.org/10.1016/j.neuropsychologia.2009.12.035

Rogalsky, C., LaCroix, A. N., Chen, K. H., Anderson, S. W., Damasio, H., Love, T., & Hickok, G. (2018). The neurobiology of agrammatic sentence comprehension: A lesion study. *Journal of Cognitive Neuroscience*, *30*(2), 234–255. https://doi.org/10.1162/jocn_a_01200

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

Schmidtke, D., Gagné, C. L., Kuperman, V., & Spalding, T. L. (2018). Language experience shapes relational knowledge of compound words. *Psychonomic Bulletin & Review*, *25*(4), 1468–1487. https://doi.org/10.3758/s13423-018-1478-x

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O. K., Dell, G. S., ... Coslett, H. B. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proceedings of the National Academy of Sciences*, *108*(20), 8520–8524. https://doi.org/10.1073/pnas.1014935108

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J., 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. https://doi.org/10.1093/cercor/bhn187

Stockall, L., Manouilidou, C., Gwilliams, L., Neophytou, K., & Marantz, A. (2019). Prefix stripping re-re-revisited: MEG investigations of morphological decomposition and recomposition. *Frontiers in Psychology*, *10*, 1964. https://doi.org/10.3389/fpsyg.2019.01964

Takashima, A., Konopka, A., Meyer, A., Hagoort, P., & Weber, K. (2020). Speaking in the brain: The interaction between words and syntax in sentence production. *Journal of Cognitive Neuroscience*, *32*(8), 1466–1483.

Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, *122*(11), 2119–2132. https://doi.org/10.1093/brain/122.11.2119

Taylor, J. S. H., Davis, M. H., & Rastle, K. (2019). Mapping visual symbols onto spoken language along the ventral visual stream. *Proceedings of the National Academy of Sciences*, *116*(36), 17723–17728. https://doi.org/10.1073/pnas.1818575116

Teige, C., Cornelissen, P. L., Mollo, G., Alam, T. R. D. J. G., McCarty, K., Smallwood, J., & Jefferies, E. (2019). Dissociations in semantic cognition: Oscillatory evidence for opposing effects of semantic control and type of semantic relation in anterior and posterior temporal cortex. *Cortex*, *120*, 308–325. https://doi.org/10.1016/j.cortex.2019.07.002

Teige, C., Mollo, G., Millman, R., Savill, N., Smallwood, J., Cornelissen, P. L., & Jefferies, E. (2018). Dynamic semantic cognition: Characterising coherent and controlled conceptual retrieval through time using magnetoencephalography and chronometric transcranial magnetic stimulation. *Cortex*, *103*, 329–349. https://doi.org/10.1016/j.cortex.2018.03.024

Thompson, H., Davey, J., Hoffman, P., Hallam, G., Kosinski, R., Howkins, S., ... Jefferies, E. (2017). Semantic control deficits impair understanding of thematic relationships more than object identity. *Neuropsychologia*, *104*, 113–125. https://doi.org/10.1016/j.neuropsychologia.2017.08.013

Tyler, L. K., Cheung, T. P., Devereux, B. J., & Clarke, A. (2013). Syntactic computations in the language network: characterizing dynamic network properties using representational similarity analysis. *Frontiers in Psychology*, *4*, 271. https://doi.org/10.3389/fpsyg.2013.00271

Tyler, L. K., Stamatakis, E. A., Bright, P., Acres, K., Abdallah, S., Rodd, J. M., & Moss, H. E. (2004). Processing objects at different levels of specificity. *Journal of Cognitive Neuroscience*, *16*(3), 351–362. https://doi.org/10.1162/089892904322926692

Van Essen, D. C. (2005). A population-average, landmark-and surface-based (PALS) atlas of human cerebral cortex. *NeuroImage*, *28*(3), 635–662. https://doi.org/10.1016/j.neuroimage.2005.06.058

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron*, *55*(1), 143–156. https://doi.org/10.1016/j.neuron.2007.05.031

Wang, J., Deng, Y., & Booth, J. R. (2019). Automatic semantic influence on early visual word recognition in the ventral occipito-temporal cortex. *Neuropsychologia*, *133*, 107188. https://doi.org/10.1016/j.neuropsychologia.2019.107188

Weisberg, J., Van Turennout, M., & Martin, A. (2007). A neural system for learning about object function. *Cerebral Cortex*, *17*(3), 513–521. https://doi.org/10.1093/cercor/bhj176

Westerlund, M., & Pylkkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia*, *57*, 59–70. https://doi.org/10.1016/j.neuropsychologia.2014.03.001

Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2011). The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cerebral Cortex*, *21*(5), 1066–1075. https://doi.org/10.1093/cercor/bhq180

Williams, A., Reddigari, S., & Pylkkänen, L. (2017). Early sensitivity of left perisylvian cortex to relationality in nouns and verbs. *Neuropsychologia*, *100*, 131–143. https://doi.org/10.1016/j.neuropsychologia.2017.04.029

Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language*, *35*(3), 434–453. https://doi.org/10.1006/jmla.1996.0024

Wisniewski, E. J., & Love, B. C. (1998). Relations versus properties in conceptual combination. *Journal of Memory and Language*, *38*(2), 177–202. https://doi.org/10.1006/jmla.1997.2550

Woolnough, O., Donos, C., Rollo, P. S., Forseth, K. J., Lakretz, Y., Crone, N. E., ... Tandon, N. (2020). Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nature Human Behaviour*, *5*, 389–398. https://doi.org/10.1038/s41562-020-00982-w

Yu, X., Bi, Y., Han, Z., Zhu, C., & Law, S. P. (2012). Neural correlates of comprehension and production of nouns and verbs in Chinese. *Brain and Language*, *122*(2), 126–131. https://doi.org/10.1016/j.bandl.2012.05.002

Yu, X., Law, S. P., Han, Z., Zhu, C., & Bi, Y. (2011). Dissociative neural correlates of semantic processing of nouns and verbs in Chinese—A language with minimal inflectional morphology. *NeuroImage*, *58*(3), 912–922. https://doi.org/10.1016/j.neuroimage.2011.06.039

Zhang, L., & Pylkkänen, L. (2015). The interplay of composition and concept specificity in the left anterior temporal lobe: An MEG study. *NeuroImage*, *111*, 228–240. https://doi.org/10.1016/j.neuroimage.2015.02.028

Ziegler, J., & Pylkkänen, L. (2016). Scalar adjectives and the temporal unfolding of semantic composition: An MEG investigation. *Neuropsychologia*, *89*, 161–171. https://doi.org/10.1016/j.neuropsychologia.2016.06.010

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.