Brain & Language 218 (2021) 104951

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

How the conceptual specificity of individual words affects incremental sentence composition: MEG evidence

Songhee Kim [a,*], Liina Pylkkänen [a,b,c]

[a] Department of Linguistics, New York University, 10 Washington Place, New York, NY 10003, USA
[b] Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA
[c] NYUAD Institute, New York University Abu Dhabi, PO Box 129188, Saadiyat Island, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Keywords:
Semantics
Semantic composition
Concepts
Magnetoencephalography
Left anterior temporal lobe
Left angular gyrus

ABSTRACT

While much research has addressed the neural basis of lexical access and the composition of lexical items into larger meanings, little is known about how the semantic properties of individual words affect composition. Research on modifier-noun combinations has, however, shown that composition related activity in the left anterior temporal lobe (LATL) is sensitive to the conceptual specificity of the composing words. Here we tested whether this pattern extends to verb-argument combinations in minimal subject-verb-object sentences. If the LATL specificity effects extend to verb-argument integration, this would suggest a general mechanism that composes not only entity concepts, but also propositions describing events. Results showed an overall similar modulation by conceptual specificity in the verb domain, suggesting a central, category-insensitive, role for the LATL as a conceptual combiner. Additionally, we saw specificity effects in the left mid-superior temporal cortex, but the angular gyrus, often hypothesized as combinatory, showed no effects of composition.

1. Introduction

When words combine with each other to form larger expressions, the representational complexity of the expression increases along both semantic and syntactic dimensions. While the neural implementation of the combinatory process is still largely unknown, at least some subroutines of semantic composition have been hypothesized to localize in left-lateral cortex both anteriorly and posteriorly, specifically in the left anterior temporal lobe (LATL) and the left angular gyrus (LAG). The role of the LATL is better understood than the role of the LAG. In MEG measurements of comprehension, the LATL shows a consistent response at around 250 ms after word onset that is increased if the word can compose with the prior word. The response is highly sensitive to the conceptual specificity of the composing items: amplitudes elicited by the second word of a phrase reflect the proportion of features contributed by the first word onto the combined feature set of the whole phrase (Pylkkänen, 2019; Westerlund & Pylkkänen, 2014; Zhang & Pylkkänen, 2015; Ziegler & Pylkkänen, 2016). Such observations have ruled out syntactic explanations of the composition effect in the LATL, suggesting instead a more conceptual role, with the early 250 ms response reflecting some type of combination of the feature sets of the composing words. However, extant evidence for this account is mostly limited to nouns and their modifiers.

To test the generalizability of the LATL's conceptual sensitivity, we investigated verb-argument composition in minimal three-word sentences, comprised of a subject, verb and object (e.g., *boys throw balls*). Conceptual specificity was varied in each position. If the LATL patterns similarly for the modification of nouns and for verb-argument composition, this would suggest a general role in category-insensitive conceptual composition. We also analyzed other regions hypothesized to participate in composition, such as the LAG, which has been associated with event concept composition (Boylan, Trueswell, & Thompson-Schill, 2015; Matchin, Liao, Gaston, & Lau, 2019). We did not however, find any evidence for a combinatory role for the LAG, though posterior temporal did show sensitivity to our manipulation.

1.1. Interplay of conceptual specificity with combinatorial operation

Conceptual specificity refers to the information or feature density of a word. It is most intuitive to understand in comparison to another word. For example, *carrot* has higher conceptual specificity than *vegetable*. Conceptual specificity has figured prominently in the study of certain clinical conditions such as semantic dementia, a neurodegenerative disease that affects specific, subordinate level concepts more severely than basic- or superordinate level concepts (Hodges, Patterson, Oxbury, &

* Corresponding author at: Medical College of Wisconsin, MEB M4678, 8701 Watertown Plank Road, Milwaukee, WI 53226, USA.
E-mail addresses: songhee.kim@nyu.edu (S. Kim), liina.pylkkanen@nyu.edu (L. Pylkkänen).

https://doi.org/10.1016/j.bandl.2021.104951
Received 20 January 2020; Received in revised form 29 March 2021; Accepted 30 March 2021
Available online 21 April 2021

Funnell, 1992; Rogers et al., 2004; Snowden, Goulding, & Neary, 1989; Warrington, 1975). As semantic dementia often exhibits atrophy in the LATL, researchers have suggested that this region may play a critical role in representing relatively specific concepts, perhaps by *binding* various features into a single conceptual representation (cf. *hub-and-spoke* model, Lambon Ralph, Sage, Jones, & Mayberry, 2010; Patterson, Nestor, & Rogers, 2007; Pobric, Jefferies, & Lambon Ralph, 2010).

Notably, complex concepts are not only found in the lexicon; they can also be assembled through composition; e.g., *red boat* creates a more specific concept than *boat*. A series of MEG studies has shown that phrasal composition also engages the LATL: in two-word composition, LATL activity increases at around 200–300 ms after the onset of the second word, compared to a non-combinatory baseline, i.e., *composition effect* (Bemis & Pylkkanen, 2011). Subsequent works have further revealed that this effect *interacts* with the conceptual specificity of input words (Westerlund & Pylkkänen, 2014; Zhang & Pylkkänen, 2015, 2018; Ziegler & Pylkkänen, 2016). That is, in noun-noun composition such as *tomato soup*, the increase in the LATL at the second noun is detected only i) when the head noun is relatively less specific (e.g., *dish* as opposed to *soup* in the head noun position) *and* ii) when the modifier is relatively more specific (e.g., *tomato* as opposed to *vegetable* in the position preceding the head noun). In other words, the largest composition effect is elicited for the combination of a more specific modifier and a less specific head noun (henceforth "specific-over-general" effect.)

This rather intricate pattern of modulation has been accounted for by assuming a feature-addition-like process for modification operation (Westerlund & Pylkkänen, 2014; Zhang & Pylkkänen, 2015; see Pylkkänen, 2019, 2020 for review). That is, when a modifier (e.g., *red*) combines with a head noun (e.g., *boat*), some features of the modifier (e.g., *redness*) are added to the featural specification of the modified word (i.e., *boat* = [vehicle, on the water, long shape, …]), resulting in a more specified representation of the head word (i.e., *red boat* = [vehicle, on the water, long shape, red, …]). Crucially, a modifier of high specificity has *more* features to add than a modifier of low specificity, and thus induces a *greater* change to the representation of the head noun. As for the head noun, the effect of feature addition will become greater when the integration is onto a less, not more, specific head noun, as a less specific head noun comes with a sparser feature space (cf., Weber's law). In other words, the LATL activity seems to be the largest when the proportional impact of the first word integrating onto the second word is the greatest. Previous research has thus suggested that the composition effect in the LATL might be driven by the *change* of conceptual specificity between concepts in isolation and in modification (Zhang & Pylkkänen, 2015).

The feature-based account explains the observed pattern by assuming LATL's susceptibility to a certain aspect of modification that could arguably be best captured in terms of feature addition, as described above. However, whether the neural correlates of conceptual specificity observed in the LATL (i.e., "specific-over-general" pattern) are strictly confined to modification has not been tested. That is, is the pattern specifically bound to the modification process? How general is the combinatorial mechanics of the LATL, which engenders to the "specific-over-general" pattern, across different types of semantic composition? In this paper, we investigated verb–argument combination to tackle this question, as it serves an ideal test case due to its divergence from noun modification both in terms of the type of combinatory operation (argument saturation as opposed to modification) *and* the type of conceptual entity being composed (events as opposed to objects).

As to the type of combinatory operation, one first needs to consider the extent the semantics of verb-argument composition may or may not overlap with that of noun modification. That is to say, is verb-argument composition a fundamentally different process from modification, which is schematized as a feature-addition process? Some theories in formal semantics have treated argument saturation and modification as two distinct modes of semantic composition (See Section 1.2 for more details). However, no clear understanding is available yet as to how the two processes might differ at the *implementation* level. When a subject noun combines with a verb (e.g., *boys* + *run*), for example, the subject does not *modify* the verb in the same way a verb modifier does (e.g., *slowly* + *run*). Nonetheless, the range of possible events denoted by the verb is *reduced* as a result of composition so that only the running events where the runners are boys are included. This is in fact analogous to the process of modification such as *tomato* + *soup*, where the addition of a modifier (*tomato*) reduces the possible range of the *soup* so that only the soup made with tomato can be referred to. That is, in both cases, conceptual composition leads to the identification of a narrower set of referents.

Given this, the process of combining a verb with its argument might have some overlap with that of modification, which can also be illustrated as a feature-addition process as follows: event concepts denoted by verbs come in with their feature space (e.g., *run* = [action, using legs, fast, have animate agent, …]) just as object concepts do, and the saturation of arguments (e.g., *boys*) fill in some open slots in the feature space (e.g., who is agent?), thus effectively reducing the possible instances of events and increasing the specificity of the resultant event space (e.g., *boys run* = [action, using legs, have animate agent, agent is boys, …]). Under this idea, the verb-argument composition could engage the LATL combinatory network in a similar way as in noun modification, and conceptual specificity manipulation might as well exert a similar effect on the composition of verb and arguments.

Apart from the type of combinatory operation, verb-argument composition also differs from noun modification in the type of concepts that compose: verb-argument composition operates in the realm of event concepts, whereas noun-modification in the realm of objects concepts. To what extent is the LATL combinatory profile generalized across different conceptual categories? Previous studies have reported the engagement of the LATL in verb phrase composition (Kim & Pylkkänen, 2019; Westerlund, Kastner, Al Kaabi, & Pylkkänen, 2015). However, the effect of conceptual specificity in a combinatory context has so far been only attested in the object domain. From the perspective of the feature-addition hypothesis, it is indeed possible that the particulars of features, such as their types or the feature geometry, might affect the LATL. For instance, object concepts are often loaded with visual attributes such as size, color, shape, as well as somatosensory ones such as weight or texture (Binder et al., 2016), but events likely are not, and might instead have motor and spatial features, along with information about core participants, (in)animacy of the participants, a default instrument if any, whether a result state is inherently implied or not, and so on. If the combinatory activity in the LATL is constrained such that its operation is subject to those particulars of features that constitute a concept, different composition mechanics might be at work in the event domain.

In sum, to better understand the nature of the conceptual specificity modulation in the LATL, we investigated the effect of conceptual specificity in the verb-argument composition. The stimuli were three-word English sentences consisting of a subject, verb, and object, with conceptual specificity being varied at each word position. This allowed us to assess the dynamic effect of conceptual specificity of one word onto the other during sentence composition.

1.2. Neural correlates of verb-argument composition

Aside from the effect of conceptual specificity, we also aimed to address the more general and fundamental question: how is verb-argument composition, or argument saturation, reflected in the brain?

In natural language, argument saturation is one of the essential modes available to combine two words, along with modification. Generative semantics (Heim & Kratzer, 1998) distinguishes these two modes as follows: in argument saturation, one word acts as a functor and takes the other word as an argument, returning a saturated meaning of the functor, with its valence reduced by the combined argument. Modification, on the other hand, does not affect the valence of the participating words. Instead, two words' meanings become intersected (e.g., *red boat* refers to things that are both red and boat). While the neural basis of modification has been characterized in some detail in past years, consistently implicating the LATL, our understanding is more elusive for argument saturation.

Extant findings on the neural correlates of argument saturation roughly divide into two categories. First, some studies have shown evidence that argument saturation affects activity in the LATL. Westerlund et al. (2015) reported increased activity in the LATL for both argument saturation and modification contexts, supporting an across-the-board nature of the LATL combinatorial operation. However, many researchers have associated inferior parietal cortex, specifically the angular gyrus, with verb-argument processing (Binder, Desai, Graves, & Conant, 2009; Bonner & Price, 2013; Boylan et al., 2015; Meltzer-Asscher, Mack, Barbieri, & Thompson, 2015; Thompson et al., 2007). In fact, the left angular gyrus (LAG) has been considered a region of combinatory semantics in prior literature, along with the LATL (Humphries, Binder, Medler, & Liebenthal, 2007; Price, Bonner, Peelle, & Grossman, 2015), and some researchers have specifically hypothesized that the LAG is a locus of combinatorial semantics *specialized for event meaning* (Boylan et al., 2015; Matchin et al., 2019). However, when the composition, event semantics and argument structure hypotheses were directly pitted against each other, the LAG was sensitive only to argument structure valence, not to composition or the eventivity of the concept (Williams, Reddigari, & Pylkkänen, 2017). In addition to the LATL and LAG, the left mid-superior temporal cortex (lmSTC) and its adjacent areas have also been implicated in the processing of thematic role information (e.g., who did what to whom) (Frankland & Greene, 2015; Wang et al., 2016; Williams et al., 2017), or more generally, argument structure processing (Matchin et al., 2019).

As we evaluate the extant evidence, it is important to keep in mind that verb-argument combination is likely a complex process. Possible subroutines include: i) postulation of possible subcategorization frames associated with a given verb (e.g., a verb *climb* would activate a total of three subcategorization frames as follows: a) no object, e.g., *John climbed*, b) a noun phrase, e.g., *John climbed **the tree***, c) a prepositional phrase, e.g., *John climbed **through the window***); ii) for any given subcategorization frame, postulation of argument slots (e.g., for the subcategorization frame c above, for example, slots for a subject (noun phrase) and for a path (prepositional phrase) need to be filled in); iii) integration of actual arguments into the slots; and iv) evaluation of the syntactic and semantic outcome of the integration (See also Thompson and Meltzer-Asscher 2014 for a similar proposal). The fact that the literature implicates various brain areas in verb-argument processing, encompassing the left inferior frontal gyrus, LAG, middle/superior temporal gyrus, and the LATL, therefore, may reflect this heterogeneous nature of the processes. As it turns out, previous studies seem to tap into different aspects of verb-argument processing with their manipulations. For example, Thompson et al. (2007) manipulated the valency of verbs (e.g., number of arguments that a verb requires), which should relate to the process of postulating subcategorization frames but not to the composition process. Similarly, Price et al. (2015) manipulated how readily the words in a noun-modifying phrase can be integrated into a meaningful phrase, which might target the semantic aspect of the combinatory operation (i.e., iv above). More direct assessment of composition signature, on the other hand, does not find the engagement of the LAG (Westerlund et al., 2015; Williams et al., 2017).

To gain an initial understanding how verbs and arguments combine, this work focuses on the contrast between composition and no composition. Expanding from the two-word composition paradigm in prior literature, we carry out our investigation at the sentence level, assessing the integration of two essential arguments, i.e., subject and object, with the verb. As in previous studies, trials with fewer words (i.e., "verb-object" condition, and "object-only" condition) were created to serve as non-combinatory baselines. With this type of incremental paradigm, we intended to capture how semantic composition unfolds over a sentence, beyond a two-word phrase.

2. Methods

2.1. Participants

Thirty-four native English speakers participated in the experiment (21 women, 18–40 years old, mean age = 27.7 years). All participants were right-handed and had normal or corrected to normal vision. Prior to the experiment, participants gave their informed consent. A total of seven participants were excluded in the analysis; one subject's recording had excessive noise, one subject fell asleep during the session, and two subjects showed poor performance (accuracy less than 70%) on the picture-matching task. Also, one subject was removed due to inaccurate marker position, another subject due to poor coregistration because of extremely high volume of hair. Finally, one subject was removed because the reference channels did not work, rendering noise-reduction impossible. Thus, a total of twenty-seven participants were included in the data analysis (17 women, 18–40 years old, mean age = 27.97 years). Four of them had structural MRIs.

2.2. Design and stimulus

We constructed a set of 42 English sentences consisting of a subject, a verb, and an object. At each position, conceptual specificity was manipulated at two levels. For example, the subject position was occupied by either a noun of high specificity (e.g., *terrorists*) or of low specificity (e.g., *criminals*). Similarly, the verb position was occupied by either a high specificity verb (e.g., *hoard*) or a low specificity verb (e.g., *store*). The object position also had a pair of nouns, one of high specificity (e.g., *gas*), the other of low specificity (e.g., *fuel*). Subject and object nouns were bare plural or mass nouns, without any preceding determiners. Verbs were presented in the bare form, which is grammatical in the present tense in English (e.g., *criminals store fuel*). This gave rise to a total of 8 (=2 × 2 × 2) conditions.

The selection of high vs. low specificity pairs was based on Wordnet (https://wordnet.princeton.edu). For nouns (i.e., subjects and objects), a high specificity noun was chosen among the hyponyms of the low specificity noun (e.g., *surgeon - physician*). A total of 42 hyponym-hypernym pairs were chosen for the object, and for the subject position, 14 pairs were chosen and repeated three times, as there were fewer pairs available that can be an agent. For the verb position, 42 high specificity-low specificity verb pairs were selected from Wordnet. For the majority of these pairs, the high specificity one was a troponym (i.e., a verb that indicates more precisely the manner of doing something that is described by the original verb) of the low specificity counterpart. Also, among the 42 verb pairs, six of them (i.e., *dislike, evaluate, examine, love, offend,* and *describe*) referred to mental state or activity, whereas the rest referred to actions (e.g., *chew*). The difference in specificity was tested via Amazon Mechanical Turk (see Section 2.4.) The list of complete stimuli is provided in Appendix.

The experiment also had non-combinatory conditions which have hashtags in place of words. That is, we created "No subject" condition by replacing the subject with three hashtags (i.e., ### verb object). And for "No verb" condition, both the subject and verb were replaced with three hashtags (i.e., ### ### object). This added six more conditions; 1x2x2 (for No subject) + 1x1x2 (for No verb), thus a total of 14 conditions. The design of the experiment is shown in the Table 1. In the remaining of this paper, we call the low specificity words as "general" and high specificity words as "specific."

2.3. Lexical properties of stimuli

Some lexical properties of the stimuli such as the frequency, number of morphemes, and length, were not perfectly controlled for across the two levels of specificity. For the subject position, the (log-transformed) frequency of general subjects was overall greater than that of specific subjects ($p < 0.01$). In addition, general subjects had fewer morphemes

Table 1
Stimulus design, varying conceptual specificity in subject, verb and object positions. Two-word and one-word conditions were included as less combinatory controls.

	Subject	Verb	Object	Example		
Condition1	GEN	GEN	GEN	*kids*	*throw*	*objects*
Condition2	GEN	GEN	SPEC	*kids*	*throw*	*balls*
Condition3	GEN	SPEC	GEN	*kids*	*toss*	*objects*
Condition4	GEN	SPEC	SPEC	*kids*	*toss*	*balls*
Condition5	SPEC	GEN	GEN	*boys*	*throw*	*objects*
Condition6	SPEC	GEN	SPEC	*boys*	*throw*	*balls*
Condition7	SPEC	SPEC	GEN	*boys*	*toss*	*objects*
Condition8	SPEC	SPEC	SPEC	*boys*	*toss*	*balls*
Condition9	###	GEN	GEN	*###*	*throw*	*objects*
Condition10	###	GEN	SPEC	*###*	*throw*	*balls*
Condition11	###	SPEC	GEN	*###*	*toss*	*objects*
Condition12	###	SPEC	SPEC	*###*	*toss*	*balls*
Condition13	###	###	GEN	*###*	*###*	*objects*
Condition14	###	###	SPEC	*###*	*###*	*balls*

than the specific subjects ($p < 0.01$). At verb position, the general and specific verbs differed both in terms of length and frequency: general verbs were shorter than specific verbs ($p < 0.02$) and had higher frequency than the specific verbs ($p < 0.01$). At object position, general and specific objects differed in length and frequency: general objects were longer (mean = 6.9) than the specific objects (mean = 6.26), and were more frequent (mean = 4.41) than the specific ones (mean = 4.21). For all of these words, frequency was taken to be the cumulative counts of all inflected forms (e.g., frequency of *devour* = the sum of the occurrences of *devour*, *devoured*, *devouring*, *devours*) in the Corpus of Contemporary American English (COCA, http://corpus.byu.edu/coca/) (Davies, 2008). Due to these unmatched lexical variables, regression analyses for single trial data were performed in the analysis stage to assess whether these differences accounted for the observed differences in the brain activity.

2.4. Amazon Mechanical Turk norming experiment

2.4.1. Specificity judgment test

For all of the general-specific word pairs used in the experiment, norming experiments were executed using Amazon Mechanical Turk. These studies aimed to check whether a specific word in a given general-specific pair was indeed judged to be more specific than its general counterpart. Turkers were asked to choose the word with a more specific meaning from a pair of words. The answer was binary, without a third option. Tests also included a few filler items with a clear difference in specificity in order to make sure the subject was paying attention to the test. Subjects who did not report themselves as native speakers of English or did not give correct answers to the filler questions were excluded in the analysis. All of the general-specific word pairs were judged as anticipated by at least 70 percent of participants.

2.4.2. Naturalness judgment test

Naturalness of each sentence was also normed to assess a potential difference in the naturalness of sentences among conditions. Turkers were randomly given a collection of 30 sentences and were asked to indicate how natural each sentence sounds, on a scale of 1 (very unnatural) to 7 (very natural). To confirm whether the Turkers were paying close attention to the stimuli, five filler items which are clearly unnatural (e.g., *koalas choke socks*) were added to each batch. A one-way ANOVA with the condition as a single factor tested found no significant effect of condition on naturalness score among the 12 conditions; $F(11, 492) = 1.33$, $p = 0.2$.

2.5. Trial structure

Each trial started with a 300 ms fixation cross. Then a subject, verb, and object appeared in the center of the screen for 300 ms, one word at a time, with intervening blank screens between words which also lasted for 300 ms. To encourage full attention on the stimulus, a picture-matching task was given at the end of every sentence: participants indicated whether a picture presented on the screen accurately described the meaning of the linguistic expression they just read by pressing a button with their left hand. Roughly half of the pictures matched the linguistic expression, and the other half did not. The screenshots of an example trial are shown in Fig. 1.

2.6. Procedure

Before the MEG recording, subjects' head shapes were digitally scanned using a Polhemus Fastscan three-dimensional laser digitizer. Participants then completed a short practice session in front of a computer outside the recording room, asking questions if they want to. Once inside the MEG machine, participants listened to a series of single tones for up to three minutes while their eyes were closed in order to collect basic brain responses to a simple audio stimulus as a sanity check. The main experiment consisted of 20 blocks, each of which contained 42 trials. The stimuli were presented via PsychoPy (Peirce, 2010). Halfway through each block, a screen appeared asking whether the subject wanted to take a break or not. The average recording time was approximately 50 min.

2.7. Pre-processing of MEG data

To align the position of MEG sensors with subjects' actual head surfaces, digitized MEG sensor positions were coregistered with the FreeSurfer average brain (CorTechs Labs Inc., Lajolla, CA). After noise-reduction of the MEG data was performed using the MEG 160 software (Kanazawa Institute of Technology), the rest of pre-processing steps and statistical analyses were performed in MNE-Python (Gramfort et al., 2014) and Eelbrain 0.27 (https://doi.org/10.5281/zenodo.1186450). First, the data were high-pass filtered at 1 Hz and low-pass filtered at 40 Hz. Bad channels were removed by visual inspection, and an independent component analysis (ICA) cleaned the data further, rejecting artifacts such as heartbeat and eye movement. The cleaned data was epoched from 200 ms before the onset of the first word to 600 ms after the onset of the third word, for a total length of 2000 ms. Epochs were rejected if their maximum amplitude exceeded 3000ft. Brain activity during the epochs was averaged per condition, and the forward solution was calculated using the Boundary Element Model (Bonnet, 1999; Mosher, Leahy, & Lewis, 1999, among others) as the source model. Finally, the inverse solution, or, the activity at the source level, was computed by calculating minimum-norm estimates (Hämäläinen & Ilmoniemi, 1994) using the free orientation of dipoles.

2.8. Data analysis

Our regions of interest (ROIs) included the LATL, i.e., the left BA 21 and left BA 38 areas, as the effect of conceptual specificity has been localized to these regions (Westerlund & Pylkkänen, 2014; Westerlund et al., 2015; Zhang & Pylkkänen, 2015). The left angular gyrus (BA 39) and the left mid-superior posterior temporal cortex (lmSTC) were also included. The lmSTC was defined as the area that is within 25 ms from the Talaraich coordinate Frankland and Greene (2015) reports, i.e., (−59, −25, 6).

2.8.1. Analysis over full sentence conditions

To assess the effect of conceptual specificity during full sentence composition, we performed a one-way ANOVA of conceptual specificity of subject, verb, and object, over the averaged time course in each ROI. Non-combinatory conditions (i.e., two-word and one-word trials) were not included in this analysis. The ANOVA was performed at the 100–600 ms of the current, and next words. For example, the effect of subject specificity was tested over the following three intervals: 100–600 ms after the onset of the subject, verb, and object. Similarly, the effect of verb specificity was tested over the same interval after the

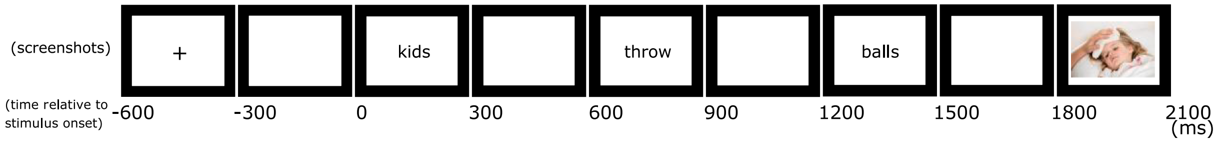


Fig. 1. Trial structure. Trials started with a fixation cross, and words appeared one at a time with intervening blank screens. All screens lasted for 300 ms.

onset of the verb and object. At the object position, in addition to the one-way ANOVA of object specificity, a three-way ANOVA (i.e., general/specific subject × general/specific verb × general/specific object) was performed as well.

All of these tests used nonparametric statistical procedures, in which the significance of an observed difference is evaluated against a simulated null distribution created by repeated random shuffling of obtained data points (Maris & Oostenveld, 2007). Specifically, after activity was averaged across all the sources within an ROI for each time sample (=1 ms), a cluster was formed if a difference between conditions was significant at the level of $p < 0.05$ (uncorr.) for at least 10 consecutive time points. Among such clusters within the analysis time window (i.e., 100–600 ms in cases above), the cluster with the biggest F-value (i.e., the sum of the F-values at each time point within the cluster interval) was chosen, and the data points within that cluster went through 10,000 random permutations, randomly shuffling the condition labels within subjects. The statistical significance of the actual data was evaluated against this null distribution, its p-value being the proportion of the permutations whose statistic is greater than the observed statistic (corrected alpha = 0.05).

2.8.2. Analysis over all conditions

To assess how conceptual specificity of words modulates the composition effect (i.e., the increase against non-combinatory baseline), one-way ANOVA was performed including non-combinatory conditions as well as full sentence conditions. Two ANOVA tests were performed. First, a one-way ANOVA of subject specificity (i.e., general/specific/no subject) was run over the time course at 100–600 ms after verb onset. Second, a one-way ANOVA of verb specificity (i.e., general/specific/no verb) was performed at the 100–600 ms time window after object onset.

2.8.3. Single trial analysis

As our stimuli were not controlled for lexical properties (i.e., number of morphemes, length, and frequency) across conditions, a two-stage regression analysis, as implemented in Eelbrain ver.0.27, was performed over single trial data to test whether any effects found were due to the unmatched lexical properties. At stage one, we fitted a multiple linear regression model for each participant's data separately; at each time point within the analysis time window (i.e., 100–600 ms after each word onset), the source estimates, averaged across all vertices in an ROI, constituted the dependent variable, and the unmatched variables were included as predictors. β-coefficient was calculated for each time point, per variable and per subject, and then one-sample *t*-test was performed to test if these values are significantly different from zero. Over the t values at each time point, clusters were formed if the t value exceeds a value that is equivalent to $p = 0.05$ (uncorrected) for more than 10 contiguous time points. Following Maris and Oostenveld (2007), the t values of each cluster were summed and then its statistical significance was evaluated against a null distribution which is computed from 10,000 permutation of the data, shuffling the predictors randomly for each permutation.

3. Results

3.1. Behavioral results

The mean accuracy across all 27 participants and all conditions was 88.31% (SD = 2.98%). The mean response time for all conditions across 27 participants was 1.15 s (SD = 0.26 s). Since the purpose of having a comprehension task was to make subjects pay full attention to the stimuli, the behavioral results were not further analyzed.

3.2. ROI results

MEG data are available at the Open Science Framework (https://osf.io/68k7g/).

3.2.1. Results over full sentence conditions

A one-way ANOVA of subject specificity (i.e., general/specific subject) over the 100–600 ms interval after the onset of a subject, verb, and object, respectively, showed two significant clusters in the lmSTC. The clusters were located at 341–391 ms after the subject onset ($p = 0.006$), and at 345–389 ms after the object onset ($p = 0.05$), respectively (Fig. 2g). In both clusters, specific subjects elicited greater activity than general subjects.

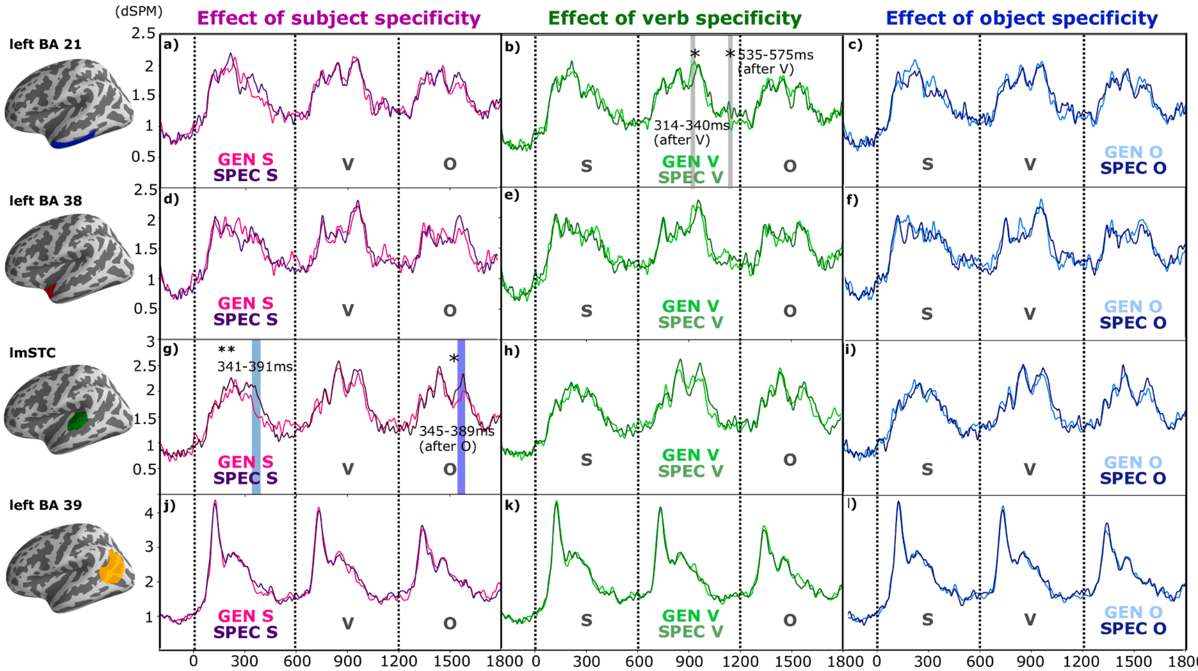
Next, a one-way ANOVA of verb specificity (i.e., general/specific verb) over 100–600 ms after the onset of a verb, and object, revealed two clusters in the left BA 21 during verb processing (Fig. 2b). One cluster was found at 314–340 ms after the verb onset ($p = 0.048$), during which a general verb elicited more activity than a specific verb. However, visual inspection of the time course suggests that this effect is due to a different latency between two conditions, not a reflection of amplitude difference. The other cluster showed a larger activity for specific verbs than general verbs ($p = 0.033$), but its timing was a bit late, i.e., 535–575 ms, compared to the timing of specificity effect in previous literature, i.e., 200–300 ms (Westerlund & Pylkkänen, 2014; Zhang & Pylkkänen, 2015). Lastly, a one-way ANOVA of object specificity (i.e., general/specific object) over the first 100–600 ms found no significant clusters (Fig. 2, blue column).

Next, a three-way ANOVA (i.e., general/specific subject × general/specific verb × general/specific object) at the end of the sentence found three clusters of interaction: i) left BA 38: verb × object, 247–276 ms, $p = 0.04$ (Fig. 3a), ii) left BA 21: subject × object, 360–401 ms, $p = 0.01$ (Fig. 3b), iii) lmSTC: subject × object, 378–419 ms, $p = 0.005$ (Fig. 3c).

The pattern of interaction was the same across all regions. That is, the combination of a specific first word followed by a general second word (i.e., specific-general) always elicited a larger activity than the combination of a general word followed by a general word (i.e., general-general). This effect of conceptual specificity was not found when the second word was a specific one. In addition, the general-specific combination elicited a larger activity than general-general combination.

In order to verify that the specific-over-general patterns, observed in verb × object interaction and subject × object interaction, are not due to a potential difference in the overall (un)predictedness of words in a sentence, we compared the Transition Probability (TP) from verb to object across conditions, using COCA corpus (https://www.english-corpora.org/coca/). T-tests showed that the TP did not differ significantly between conditions; i) genV-genO vs. genV-specO, $t(60.971) = -0.36$, $p = 0.72$, ii) genV-genO vs. specV-genO, $t(60.214) = -1.18$, $p = 0.24$, iii) genV-genO vs. specV-specO, $t(41.889) = -1.06$, $p = 0.30$, iv) genV-specO vs. specV-genO, $t(81.96) = -0.66$, $p = 0.51$, v) genV-specO vs specV-specO, $t(44.415) = -0.96$, $p = 0.34$, vi) specV-genO vs., specV-specO, $t(44.567) = -0.77$, $p = 0.44$.

For subject × object interaction, direct calculation of subject-to-object TP was not possible, as nouns do not appear adjacently in English. Instead, we employed Latent Semantic Analysis (http://lsa.colorado.edu/), which provides a measure of semantic similarity between



(GEN S = General Subject, SPEC S = Specific Subject, GEN V = General Verb, SPEC V = Specific Verb, GEN O = General O, SPEC O = Specific O)
*: $p < 0.05$, **: $p < 0.01$ (uncorrected)

Fig. 2. Main effects of subject, verb, and object specificity during sentence composition. The effect of subject specificity (first column) was found in the lmSTC, where the specific subject elicited more activity than the general subject during the subject presentation. This contrast appeared at the object presentation again. The effect of verb specificity (second column) was found in the left BA 21, where the effect was found, again, during the verb presentation. The effect of object specificity was not found (third column).

words or texts in LSA semantic space. We calculated the LSA similarity value between a subject noun and an object noun, and performed t-tests to see if the values differed significantly in any two conditions. In our stimuli, a total of two subject-object pairs had missing values in LSA (namely, officers-*felons* in genS-genO condition, and *detectives-felons* in specS-genO condition). The semantic similarity between subject and noun as measured by LSA was not significantly different across conditions; i) genS-genO vs. genS-specO, $t(77.263) = -0.51$, $p = 0.61$, ii) genS-genO vs. specS-genO, $t(76.247) = 0.09$, $p = 0.92$, iii) genS-genO vs. specS-specO, $t(80.511) = -0.54$, $p = 0.58$, iv) genS-specO vs. specS-genO, $t(68.877) = 0.63$, $p = 0.53$, v) genS-specO vs. specS-specO, $t(80.312) = 0.0063$, $p = 0.99$, vi) specS-genO vs. specS-specO, $t(74{,}568) = -0.689$, $p = 0.49$.

3.2.2. Results over all trials

Subject-verb integration

A one-way ANOVA of subject specificity over the verb found two clusters in the LATL: i) left BA 38, 313–374 ms ($p = 0.0021$) (Fig. 4a), ii) left BA 21, 327–376 ms ($p = 0.0055$) (Fig. 4c). In these clusters, both the general and specific subjects elicited more activity than a no-subject condition, without a reliable difference between the first two.

The same test over the lmSTC showed a significant ($p = 0.027$) effect of subject specificity at 332–364 ms, during which i) both the general and specific subjects elicited a significantly larger activity than no subject condition, and ii) the size of increase was larger for the specific subject than for the general subject (Fig. 4b). This pattern is reminiscent of the pattern found in the left BA 38 in Zhang and Pylkkänen (2015), in which the composition effect elicited at the noun was larger when the preceding modifier was a more specific one. Lastly, no effect was found in the LAG (Fig. 4d).

Verb – object composition

A one-way ANOVA of verb specificity (i.e., general/specific/no verb) was performed over the 100–600 ms of the object presentation. No significant clusters were found in any of the ROIs. An additional spatiotemporal test was performed over the whole left hemisphere searching for a cluster showing an effect of verb-onto-object integration, but no significant cluster was found.

3.2.3. Single trial analysis results

Since some lexical variables in the stimuli, such as length, frequency and number of morphemes, were not perfectly matched for across conditions, a two-stage multiple linear regression was performed over the single trial data to address whether any of the observed differences was attributable to these uncontrolled-for lexical variables (see 2.8.3 for details). No significant cluster was found when the ROI time courses were regressed on these lexical variables of the subject, verb, and object.

4. Discussion

The primary goal of this study was to understand how conceptual specificity modulates the incremental composition of a simple sentence consisting of a verb and its arguments. In particular, we asked whether the characteristic modulation of conceptual specificity observed in noun modification, such that the greatest composition effect is elicited for the combination of a specific word followed by a general word (i.e., "specific-over-general" pattern), extends to other types of composition. Our results show that conceptual specificity modulates LATL activity in verb-argument contexts as well, with overall similar, but not exactly the same, pattern. While the "specific-over-general" pattern does extend to verb-argument composition, such modulations were all found at the *end* of the sentence. In other words, at the verb, which was located in the middle of the sentence, no modulation by conceptual specificity of a preceding word was detected, unlike in the noun domain. The lmSTC, on the other hand, showed an effect of conceptual specificity at the verb, with its pattern more closely replicating the results in the noun domain, i.e., a bigger composition effect at the verb appeared when a more

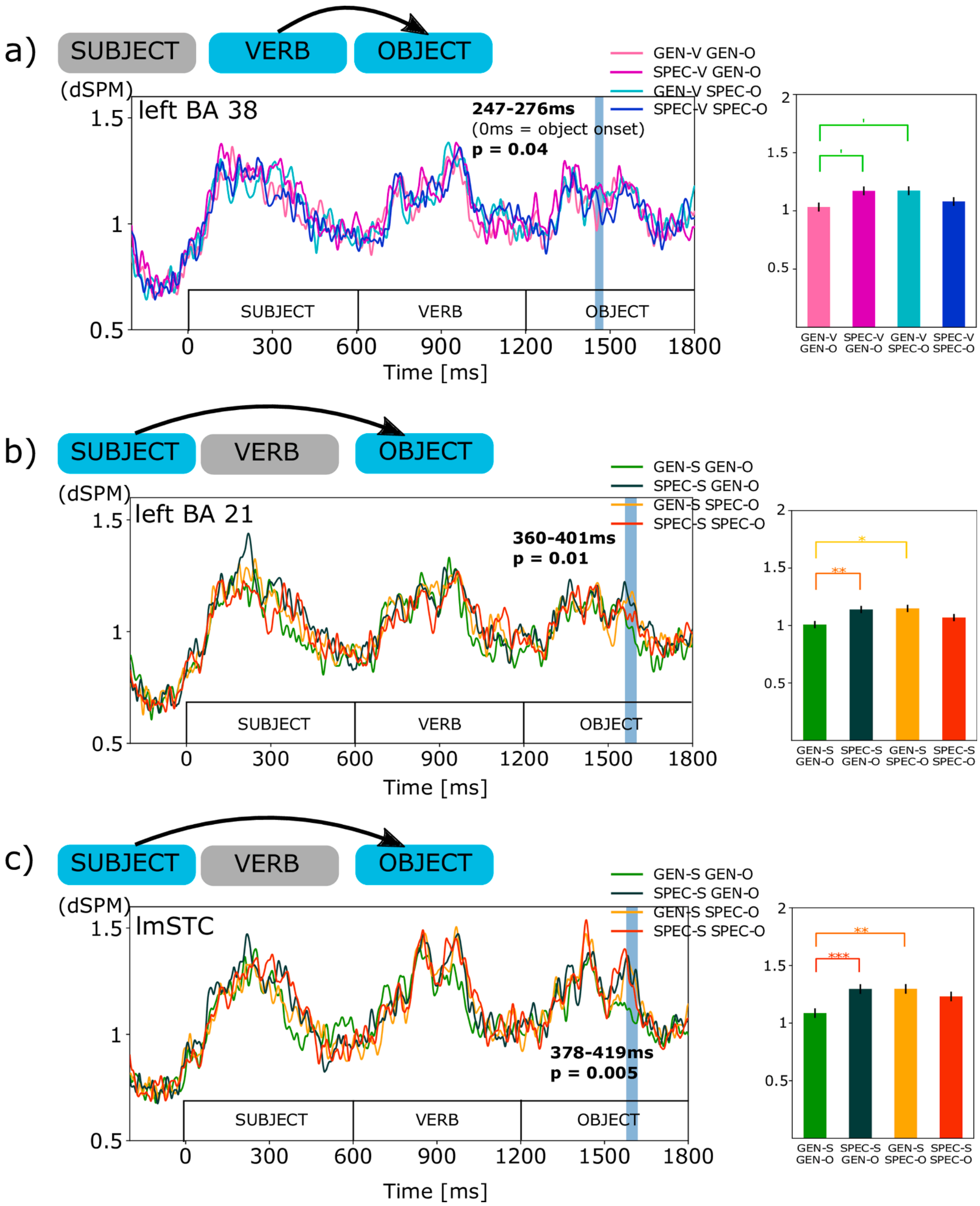


Fig. 3. Interaction effects among subject, verb, and object specificity at the presentation of object. An interaction between verb and object specificity was found in the left BA 38 (a), and an interaction effect between subject and object specificity was found in the left BA 21 (b) and in the lmSTC (c). All of the interactions followed the pattern such that the specific first word + general second word elicited a larger activity than the general first word + general second word, which is reminiscent of the "specific-over-general" pattern reported in the noun domain.

specific subject is integrating onto it. Lastly, the LAG did not show any evidence of engaging with a combinatory computation in the verb-argument composition.

Some cautions are required in interpreting the data of our study, however. While a similar pattern of specificity interaction was found in the same area as previously reported in the noun modification, it does not necessarily mean that any similarly-timed effect in the LATL reflects a combinatorial operation (i.e., "reverse inference," (Poldrack, 2006)). This issue becomes important especially because we had some conditions between which some psycholinguistic variables were not perfectly matched. This means that we cannot perfectly rule out the possibility that some of our effects might be attributable to the unmatched nuisance variables.

4.1. Sensitivity of anterior and middle temporal cortex to conceptual specificity during verb-argument composition

The results indicate that the characteristic pattern of conceptual specificity modulation, reported in the noun domain, may not be unique to the modification operation. All of the interaction effects found in the current study exhibited the same pattern: i) when the second word was a general one, a specific first word elicited a larger activity than a general first word did, and ii) when the second word was a specific one, no effect of specificity of the preceding word was present.

Our LATL results, however, show that modulation by conceptual specificity in the verbal domain is not exactly the same with results from noun domain. For example, the subject-to-verb integration in the current study did *not* show an effect of subject specificity onto the next word. That is, the composition effect elicited at the verb was not affected by the conceptual specificity of the subject argument. The "specific-over-general" pattern, instead, appeared at the end of the sentence, i.e., object time (Fig. 3). This indicates a possibility that in the LATL, the conceptual composition takes place at the end of a phrase (or sentence), whereas structural composition perhaps takes place in a more incremental way. Another region of interest, the lmSTC, on the other hand, showed a different pattern, such that the effect of conceptual specificity was shown not only at the end of the sentence, but also in the middle, namely

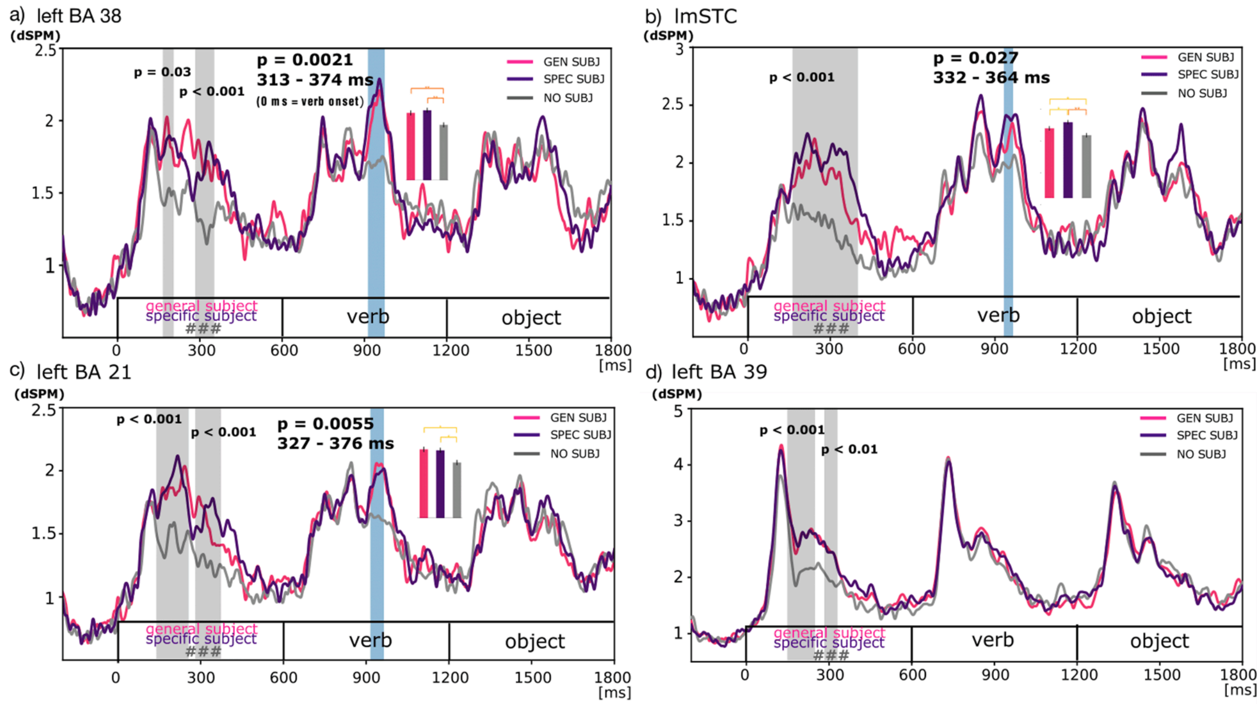


Fig.4. Subject-to-verb composition affects all the ROIs except the LAG. At the first (i.e., subject) position, conditions with a subject noun elicited more activity than the no-subject condition in all four of the ROIs. This contrast between + subject and -subject reemerged at the following word (i.e., verb), signaling the potential integration of the subject onto the verb. This subject-to-verb composition effect was found in all the ROIs except the left BA 39. Unlike in the noun domain, the subject-to-verb composition effect in the left BA 38 and 21 was not modulated by the conceptual specificity of the preceding noun. The lmSTC, on the other hand, showed a more semantically-nuanced pattern such that the specific subject elicited a larger increase against the no-subject condition compared to the general subject (Fig. *b*, at verb).

at the verb; i.e., the effect of subject-to-verb composition was greater for a specific subject than for a general subject.

The hypothesis that the conceptual composition in the LATL may be delayed until the phrasal or sentential boundary whereas structural composition can happen more incrementally is, in fact, consistent with prior findings in literature. In studies adopting two-word paradigm, the effect of conceptual specificity was reported at the second word, which was a phrasal boundary (Westerlund & Pylkkänen, 2014; Zhang & Pylkkänen, 2015; Ziegler & Pylkkänen, 2016). When the paradigm involved more than two words, the effect of conceptual specificity was reported at the end of the sentence, i.e., in Zhang and Pylkkänen (2018), the effect of conceptual specificity of a noun phrase appeared during the verb which was located at the end of the sentence.

What does the pervasiveness of the "specific-over-general" pattern across the domain tell us about the nature of the LATL conceptual combinatory operation? Given that this pattern manifests with a range of different combinations, i.e., verb + object, subject + object, and modifier + noun as reported in previous studies, it might be tapping into a word-by-word composition process that underlies potentially all types of semantic composition. Particularly, the fact that the combination of a subject and an object, which do not stand in any grammatical relation where one can *combine* with the other, elicits the same pattern opens up a possibility that the LATL can perhaps engage itself with much more extended types of semantic composition than previously thought. Not only is it engaged with conceptual enrichment of an object concept (noun modification), it might compose an event concept (verb + object), and presumably an aspect of a propositional meaning which involves event participants (subject + object). The precise nature of those conceptual representation, and how the underlying mechanism at the LATL uniformly contributes to the computations of them need to be further studied in the future study.

Lastly, it should be noted that due to the unmatched variables across some conditions, namely length, number of morphemes, and frequency, we cannot fully exclude the possibility that some of the effects might be driven by those nuisance variables.

4.2. Distinct profiles of the LATL, lmSTC, and LAG

Although the effect of conceptual specificity was found both in the LATL, and lmSTC, the specifics of the modulation suggest the lmSTC and LATL might be involved with different aspects of verb-argument composition. First, while the LATL exhibited a variety of interaction effects involving all three words in the sentence, i.e., verb × object, subject × object, the lmSTC exhibited an interaction involving only the arguments, i.e., subject × object. In fact, no effect involving the verb specificity was found in the lmSTC. This indicates that the lmSTC might be specialized in argument processing, not necessarily engaging with the whole event concept, of which the argument is a part of. On the other hand, the LATL might be engaged with the composition of the whole event concept, which involves both the predicate and arguments. Finally, the LAG did not show any effect of conceptual specificity. Crucially, this area did not exhibit the basic composition effect, either, i.e., subject-onto-verb effect, although the presence of a subject argument did make a difference during its own presentation (Fig. 4d). This suggests that the LAG does not perform a computation of combinatory nature in the verb-argument composition context, but rather focuses on the current word.

4.3. Absence of verb-to-object composition effects

While we found a clear effect of subject-to-verb composition in our study, the effect of verb-to-object composition was not detected; i.e., no reliable contrast between ### verb object and ### ### object conditions was found, contra Westerlund et al.'s (2015) results. What might account for the absence of verb-to-object composition effect in current study? One hypothesis is that our experimental paradigm was not ideal for measuring verb-to-object integration. Due to the restriction in the design that every trial has three slots, the two-word and one-word conditions started with a nonword, namely three hashtags. Potentially, the presence of hashtags in the beginning could have affected the two-word trials (i.e., ### verb object) so that it is processed differently from a simple verb object sequence. (For example, the presence of ### in the first screen could have made people less engaged with combinatory operation, since for a half of the time they see hashtags in the beginning, the second one was also hashtags, i.e., one-word condition). In fact, given the absence of explicit instruction on how to interpret the hashtag-containing stimuli, there could be inter-subject variability in the processing of hashtag-initial conditions. On top of this, the presence of two consecutive hashtags in the object-only condition (i.e., ### ### object) could have created a less natural stimulus, preventing this condition from serving as an ideal baseline.

Another possibility has to do with a potential effect of the use of bare verb form in the current experiment. That is, unlike in Westerlund et al. (2015) where a third person singular form (e.g., *eats meat*) was used, we used a bare form of verb (e.g., *eat meat*). Crucially, in a verb + object sequence, the use of a bare form could be read as an imperative sentence, the meaning of which could be encoded differently than a verb phrase in a declarative sentence.

5. Conclusion

This paper discussed how conceptual specificity modulates the incremental composition of a sentence consisting of a verb and its arguments. As in the noun domain, verb-argument composition was also modulated by conceptual specificity, exhibiting an overall similar pattern of modulation in the LATL as observed in noun-modification. Our results found the engagement of the lmSTC in this context as well, which seems to be selectively sensitive to the specificity of arguments, not predicates. We did not find any evidence for the engagement of the LAG in the verb-argument combinatory computation. Together, our results support the claim that the LATL is the primary locus of combinatory operation across different types of composition contexts.

Funding

This study was supported by grants from the National Science Foundation (BCS-1221723 to LP, BCS-1823913 to SK); and the NYUAD Institute (G1001 to LP).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Christian Brodbeck for providing help on technical issues.

Statement of significance

This study advances our understanding of the brain's semantic combinatory network. We elucidate the computational profile of the LATL in the verb phrase composition, which supports its category-insensitive nature as a conceptual combiner. We also delineate the functional division among core brain regions in event concept composition.

Appendix. The list of stimuli

SET	SUBJECT		VERB		OBJECT	
	General	Specific	General	Specific	General	Specific
1	physicians	surgeons	alter	revise	writings	reports
2	physicians	surgeons	apply	spread	medicine	ointment
3	clerks	receptionists	organize	label	books	magazines
4	soldiers	sergeants	break	smash	devices	phones
5	protesters	rioters	burn	scorch	chairs	stools
6	kids	boys	chew	crunch	vegetables	carrots
7	employees	waiters	clean	sweep	floors	decks
8	performers	dancers	cook	fry	fish	salmon
9	cooks	chefs	cut	slice	meat	turkey
10	writers	journalists	dislike	despise	professionals	lawyers
11	artists	painters	draw	outline	landscapes	clouds
12	politicians	governors	drink	sip	liquor	whiskey
13	clerks	receptionists	eat	devour	snacks	chips
14	officers	detectives	enter	invade	residences	mansions
15	artists	painters	evaluate	criticize	artworks	paintings
16	officers	detectives	examine	inspect	stuff	luggage
17	writers	journalists	follow	chase	thieves	muggers
18	criminals	terrorists	grab	clutch	belongings	purses
19	kids	boys	grip	squeeze	handles	knobs
20	employees	waiters	heat	boil	liquid	water
21	writers	journalists	hide	bury	documents	letters
22	cooks	chefs	hit	slap	bugs	flies
23	officers	detectives	hurt	cripple	felons	murderers
24	workers	farmers	lift	heave	tools	hooks
25	criminals	terrorists	lock	bolt	cabinets	safes
26	artists	painters	love	adore	plants	flowers
27	workers	farmers	modify	renovate	buildings	cabins
28	politicians	governors	offend	humiliate	opponents	enemies
29	soldiers	sergeants	release	acquit	minors	teens
30	clerks	receptionists	remove	erase	marks	stains
31	protesters	rioters	ruin	vandalize	properties	cars
32	protesters	rioters	search	comb	places	offices

(*continued on next page*)

(*continued*)

SET	SUBJECT		VERB		OBJECT	
	General	Specific	General	Specific	General	Specific
33	workers	farmers	sell	peddle	fruits	apples
34	performers	dancers	sew	stitch	tops	blouses
35	cooks	chefs	stir	whip	eggs	yolks
36	criminals	terrorists	store	hoard	fuel	gas
37	physicians	surgeons	support	fund	organizations	hospitals
38	performers	dancers	taste	savor	desserts	cakes
39	kids	boys	throw	toss	objects	balls
40	soldiers	sergeants	dig	bore	holes	tunnels
41	employees	waiters	wet	soak	fabric	linen
42	politicians	governors	describe	clarify	ideas	plans

References

Bemis, D., & Pylkkanen, 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.

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology, 33*(3–4), 130–174. https://doi.org/10.1080/02643294.2016.1147426.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex, 19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055.

Bonner, M. F., & Price, A. R. (2013). Where Is the Anterior Temporal Lobe and What Does It Do? *Journal of Neuroscience, 33*(10), 4213–4215. https://doi.org/10.1523/JNEUROSCI.0041-13.2013.

Bonnet, M. (1999). Boundary Integral Equation Methods for Solids and Fluids. *Meccanica, 34*(4), 301–302. https://doi.org/10.1023/A:1004795120236.

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.

Davies, M. (2008). The Corpus of Contemporary American English (COCA). Available online at https://www.english-corpora.org/coca/.

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.

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

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing, 32*(1), 35–42. https://doi.org/10.1007/BF02512476.

Heim, I., & Kratzer, A. (1998). Semantics in Generative Grammar. In *Semantics in Generative Grammar* (Vol. 2, Issues 1–4). https://doi.org/10.2307/417746.

Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain, 115*(6), 1783–1806. https://doi.org/10.1093/brain/115.6.1783.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: An fMRI study. *NeuroImage, 36*(3), 924–932. https://doi.org/10.1016/j.neuroimage.2007.03.059.

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*(June 2018), 18–27. https://doi.org/10.1016/j.bandl.2018.11.003.

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

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

Matchin, W., Liao, C.-H., Gaston, P., & Lau, E. (2019). Same words, different structures: An fMRI investigation of argument relations and the angular gyrus. *Neuropsychologia, 125*, 116–128. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2019.01.019.

Meltzer-Asscher, A., Mack, J. E., Barbieri, E., & Thompson, C. K. (2015). How the brain processes different dimensions of argument structure complexity: Evidence from fMRI. *Brain and Language, 142*, 65–75. https://doi.org/10.1016/J.BANDL.2014.12.005.

Mosher, J. C., Leahy, R. M., & Lewis, P. S. (1999). EEG and MEG: Forward solutions for inverse methods. *IEEE Transactions on Biomedical Engineering, 46*(3), 245–259. https://doi.org/10.1109/10.748978.

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.

Peirce, J. (2010). PsychoPy - Psychology software for Python. *Integration The Vlsi Journal*.

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010). Amodal semantic representations depend on both anterior temporal lobes: Evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia, 48*(5), 1336–1342. https://doi.org/10.1016/j.neuropsychologia.2009.12.036.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004.

Price, A. R., Bonner, M. F., Peelle, J. E., & Grossman, M. (2015). Converging Evidence for the Neuroanatomic Basis of Combinatorial Semantics in the Angular Gyrus. *Journal of Neuroscience, 35*(7), 3276–3284. https://doi.org/10.1523/JNEUROSCI.3446-14.2015.

Pylkkänen, L. (2019). The neural basis of combinatory syntax and semantics. *Science, 366*(6461), 62–66. https://doi.org/10.1126/science.aax0050.

Pylkkänen, L. (2020). Neural basis of basic composition: What we have learned from the red-boat studies and their extensions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 375*(1791), 20190299. https://doi.org/10.1098/rstb.2019.0299.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and Deterioration of Semantic Memory: A Neuropsychological and Computational Investigation. *Psychological Review, 111*(1), 205–235. https://doi.org/10.1037/0033-295X.111.1.205.

Snowden, J., Goulding, P., & Neary, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioural Neurology*. http://doi.apa.org/psycinfo/1992-02247-001.

Thompson, C. K., Bonakdarpour, B., Fix, S. C., Blumenfeld, H. K., Parrish, T. B., Gitelman, D. R., & Mesulam, M. M. (2007). Neural correlates of verb argument structure processing. *Journal of Cognitive Neuroscience, 19*, 1753–1767. https://doi.org/10.1162/jocn.2007.19.11.1753.

Wang, J., Cherkassky, V. L., Yang, Y., Chang, K. K., Vargas, R., Diana, N., & Just, M. A. (2016). Identifying thematic roles from neural representations measured by functional magnetic resonance imaging. *Cognitive Neuropsychology, 33*(3–4), 257–264. https://doi.org/10.1080/02643294.2016.1182480.

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology, 27*(4), 635–657. https://doi.org/10.1080/14640747508400525.

Westerlund, M., Kastner, I., Al Kaabi, M., & Pylkkänen, L. (2015). The LATL as locus of composition: MEG evidence from English and Arabic. *Brain and Language, 141*, 124–134. https://doi.org/10.1016/j.bandl.2014.12.003.

Westerlund, M., & Pylkkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia, 57*(1), 59–70. https://doi.org/10.1016/j.neuropsychologia.2014.03.001.

Williams, A., Reddigari, S., & Pylkkänen, L. (2017). Early sensitivity of left perisylvian cortex to relationality in nouns and verbs. *Neuropsychologia, 100*(November 2016), 131–143. https://doi.org/10.1016/j.neuropsychologia.2017.04.029.

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.

Zhang, L., & Pylkkänen, L. (2018). Semantic composition of sentences word by word: MEG evidence for shared processing of conceptual and logical elements. *Neuropsychologia, 119*, 392–404. https://doi.org/10.1016/j.neuropsychologia.2018.08.016.

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.