Brain & Language 141 (2015) 124–134

Contents lists available at ScienceDirect

Brain & Language

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

The LATL as locus of composition: MEG evidence from English and Arabic

Masha Westerlund [a,*], Itamar Kastner [b], Meera Al Kaabi [b], Liina Pylkkänen [a,b,c]

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

A R T I C L E I N F O

Article history:
Accepted 14 December 2014

Keywords:
Left anterior temporal lobe
Semantics
Composition
MEG

A B S T R A C T

Neurolinguistic investigations into the processing of structured sentences as well as simple adjective-noun phrases point to the left anterior temporal lobe (LATL) as a leading candidate for basic linguistic composition. Here, we characterized the combinatory profile of the LATL over a variety of syntactic and semantic environments, and across two languages, English and Arabic. The contribution of the LATL was investigated across two types of composition: the optional modification of a predicate (modification) and the satisfaction of a predicate's argument position (argument saturation). Target words were presented during MEG recordings, either in combinatory contexts (e.g. "eats meat") or in non-combinatory contexts (preceded by an unpronounceable consonant string, e.g. "xqkr meat"). Across both languages, the LATL showed increased responses to words in combinatory contexts, an effect that was robust to composition type and word order. Together with related findings, these results solidify the role of the LATL in basic semantic composition.

1. Introduction

Using language to communicate requires a productive way of composing the basic elements of the lexicon into complex representations of novel ideas. While this combinatory ability is at the core of the language faculty, little is understood about the internal architecture of composition; namely, what its underlying computations are and how they are neurally instantiated. Many studies that have investigated the neural architecture of the composition have implicated the left anterior temporal lobe (LATL) as a prime candidate for a role in combinatory processes (Bemis & Pylkkänen, 2011, 2013; Brennan et al., 2010; Dronkers & Wilkins, 2004; Friederici, Meyer, & Von Cramon, 2000; Humphries, Binder, Medler, & Liebenthal, 2006, 2007; Mazoyer et al., 1993; Pallier, Devauchelle, & Dehaene, 2011; Rogalsky & Hickok, 2009; Stowe et al., 1998; Xu, Kemeny, Park, Frattali, & Braun, 2005). Specifically, the bulk of the evidence for the LATL's involvement in composition has emerged from hemodynamic comparisons of structured or meaningful sentences as compared to meaningless sentences or to word lists (Friederici et al., 2000; Humphries et al., 2006, 2007; Mazoyer et al., 1993; Pallier et al., 2011; Rogalsky & Hickok, 2009; Stowe et al., 1998), which consistently show increased LATL activity for structured stimuli. Of course, contrasting sentential material to word lists or jabberwocky sentences could engage many processes other than composition per se, such as thematic role assignment or reference resolution. To narrow down the specific role that the LATL plays in interpreting structured sentences, it is crucial to construct more minimal comparisons that vary only the presence or absence of composition.

Recent work has done precisely this, focusing on a very basic combinatory operation—the composition of an adjective and a noun to create a modified noun phrase. Bemis and Pylkkänen (2011, 2013) designed a minimal paradigm in which the same noun is presented twice: in a combinatory context (in a simple two-word phrase, preceded by an adjective, such as 'red boat') and in a non-combinatory context (preceded by an unpronounceable consonant string, e.g. 'xkq boat' or in a list, e.g. 'cup, boat'). They found increased activity in the LATL for a comparison between the modified noun and the noun in isolation. Importantly, there was no difference between the noun in a list and the noun in isolation. This pattern suggests that LATL activity is specifically engaged by composition itself, rather than a high-level sentential process or a low-level difference between the number of lexical items in the stimulus. These results have been replicated in both listening and reading (Bemis & Pylkkänen, 2013) as well as in production (Pylkkänen, Bemis, & Blanco Elorrieta, 2014), suggesting they are not modality-specific. Further, similar results have been

* Corresponding author.
E-mail address: masha.westerlund@nyu.edu (M. Westerlund).

http://dx.doi.org/10.1016/j.bandl.2014.12.003
0093-934X/

obtained by Baron, Thompson-Schill, Weber, and Osherson (2010), Baron and Osherson (2011), who showed that the LATL is engaged in conceptual combination across and within words.

An important question regarding the contribution of the LATL to composition is whether its participation is syntactic or semantic in nature. The original Bemis and Pylkkänen (2011, 2013) studies do not speak to this directly: in their design, syntactic and semantic composition co-varied. However, more recent results have shown that MEG activity that localizes to the LATL is sensitive to the conceptual specificity of the composed items within syntactically parallel expressions (i.e. adjective-noun phrases), a finding that cannot be explained in terms of syntax (Westerlund & Pylkkänen, 2014). Instead this result connects the LATL composition effects to a large literature on the role of the LATL in semantic memory and conceptual processing, suggesting that the LATL may be more specifically involved in the composition of complex concepts (a type of process that is often referred to as 'conceptual combination'), as opposed to very general syntactic or semantic composition. This kind of hypothesis is also supported by a recent MEG investigation of noun phrase production, where adjectival modification (*blue cups*) but not numeral quantification (*two cups*) engaged the LATL, the former involving conceptual combination and the latter arguably not (Del Prato & Pylkkanen, 2014). Relatedly, activity in left anterior temporal cortex has shown both N400-type semantic effects (Halgren et al., 2002; Lau, Gramfort, Hämäläinen, & Kuperberg, 2013; Nobre & McCarthy, 1995) as well as increased amplitudes for conceptually contentful as opposed to grammatical words (Nobre & Mccarthy, 1995). Further, neuropsychological data show that neither semantic dementia patients with LATL atrophy nor patients with LATL resections exhibit profound grammatical deficits (Gorno-Tempini et al., 2004; Hodges, Patterson, Oxbury, & Funnell, 1992; Kho et al., 2008; Noppeney, Price, Duncan, & Koepp, 2005; Wilson, Galantucci, Tartaglia, & Gorno-Tempini, 2012), conforming to the hypothesis that the LATL is responsible for more conceptual as opposed to grammatical aspects of composition. Traditionally, the most cited arguments for syntactic processing in the LATL have been its sensitivity to sentential structure even in pseudoword sentences (Humphries, Love, Swinney, & Hickok, 2005; Humphries et al., 2006) and the correlation between LATL damage and types of agrammaticism in large scale lesion-symptom mapping studies (Dronkers & Wilkins, 2004; Magnusdottir et al., 2013). The extent to which pseudoword sentences are void of meaning is however, not obvious (Pylkkänen, Brennan, & Bemis, 2011; Westerlund & Pylkkänen, 2014), nor have all studies examining them replicated the finding (e.g., Pallier et al., 2011). The behavioral contrast between semantic dementia vs. stroke patients is substantially more complicated to understand and presumably interacts with the sizes of the lesions overall. Collectively, however, the currently available data on the LATL are more easily explained in terms of conceptual-semantic, as opposed to syntactic, processing.

So far, experiments investigating the basic composition of conceptually contentful items have narrowly focused on adjectival modification, in part due to the ease with which it lends itself to a simple phrase-image matching task (Bemis & Pylkkänen, 2011, 2013). The literature on conceptual combination has also focused almost exclusively on noun phrases (Costello & Keane, 2000; Hampton, 1997; Medin & Shoben, 1988; Murphy, 1990; Wisniewski, 1997). As such, it is unclear exactly how generalizable the LATL effect may be to other types of composed phrases. Therefore, it remains difficult to connect findings from the minimal composition paradigm to the sentence-level literature. The goal of the current investigation was to significantly enrich the linguistic environments investigated using the minimal composition paradigm, to assess the generality of the effect.

As a first step towards determining the robustness of the combinatory response to variations in semantic environments, we investigated a broad division within the theoretical literature of composition into two main types: the satisfaction of a predicate's argument position *(argument saturation*), and the optional modification of a predicate (*modification*) (Heim & Kratzer, 1998). Argument saturation or, more formally, "function application" (Montague, 1974) is in some sense the central engine of semantic composition. Predicates require arguments in order to take part in well-formed expressions; thus, the construction of every well-formed sentence requires some amount of argument saturation. For example, in the short phrase 'kiss Fido,' the verb 'kiss' is a predicate that takes the direct object, 'Fido,' as its internal argument. Composing the predicate with its argument(s) reduces, or *saturates*, its argument requirements, allowing it to be interpreted. Some verbs cannot be part of a well-formed expression unless their argument is saturated, as seen by the ill-formedness of an expression such as 'Billy kissed,' where the direct object of 'kissed' is missing. Although verbs and their arguments are perhaps the most canonical example of argument saturation, this operation also composes prepositional phrases, determiner phrases, and the many functional layers of fully inflected sentences (Heim & Kratzer, 1998).

While many well-formed sentences plausibly compose via argument saturation alone, natural language also contains optional elements, or *modifiers*, that are used to enrich the meaning of an expression rather than satisfy any requirements of interpretability. For example, although the sentence 'the cat purred' is a perfectly grammatical description of an event, there may be situations where sentences such as 'the cat purred loudly' or 'the cat purred quietly' may be more communicatively appropriate. Similarly, the combination of an adjective such as 'black' with a noun such as 'dog' simply enriches the lexical meaning of 'dog' but is not required as part of the meaning of the word. In the psychology literature, this type of operation, in which complex concepts are built from simpler ones, is typically called *conceptual combination* (Osherson & Smith, 1981; Smith & Osherson, 1984; Smith, Osherson, Rips, & Keane, 1988). In the linguistic literature, Heim and Kratzer (1998) propose that many modification structures are interpreted via a rule entitled "predicate modification," a composition rule distinct from function application.

Current evidence showing an engagement of the LATL in adjective-noun phrases does not allow us to distinguish between a narrow hypothesis, in which the LATL is selectively engaged in modification operations, and a more general hypothesis, in which the LATL is engaged by conceptual composition more generally. While a modification-specific account would be surprising in light of the hemodynamic literature, as it would entail that the larger LATL amplitudes for sentences would be solely carried by the composition of modifiers, it remains a possible interpretation of the data. Disambiguating these hypotheses would significantly advance our understanding of the combinatory profile of the LATL; therefore, we compared LATL responses to examples of argument saturation and of modification across two different languages (English and Arabic) in two separate experiments. In the first experiment, we presented English words in six different combinatory contexts, three of which were instances of modification and three of which were instances of argument saturation. By including three examples of each rule type, we ensured that our design captured a distinction between these two composition types that abstracted over variations in word class and phrase type. We conducted a second experiment in Arabic with the aim of extending the cross-linguistic profile of the combinatory effect. In designing these experiments, we assumed that results from the word list condition used by Bemis and Pylkkänen (2011, 2013) have satisfactorily showed that the LATL increases for two-word combinatory stimuli are driven by composition and not by extra lexical processing attributable to seeing two words instead of one. Thus, we forewent the list condition in the interest of keeping the experiments short.

2. Methods

2.1. Design and materials

2.1.1. English

In order to manipulate composition without introducing sentence-level processing, we adapted the minimal paradigm introduced in Bemis and Pylkkänen (2011). Words were presented either in non-combinatory contexts (preceded by a consonant string), or in minimal combinatory contexts, in which the word was preceded by another word that composes with it. Our main motivation for using a single word preceded by a consonant string as a control was that it is nearly impossible to construct a stimulus that consists of two meaningful words, but that participants will not attempt to comprehend as a novel concept (for example, two words that represent a novel compound, such as 'cup boat', may still be interpreted by a reader as a boat that carries cups). Crucially, Bemis and Pylkkänen (2011, 2013) showed in three experiments that two words do not elicit more LATL activity than a single word control unless the words are composed into a meaningful unit, which suggests that the difference in the number of words between the conditions does not drive LATL effects.

We included two composition types: modification and argument saturation. As mentioned previously, each composition type was subdivided into three syntactically distinct instances of that composition mode (see Fig. 1 for design and examples), but data collected in the New York lab were generally too noisy to allow for analyses of individual subconditions. Instead, to increase our power, our analysis focused on composition types collapsed across these different subtypes. For modification, the subtypes were adjectival modification of nouns, adverbial modification of verbs, and adverbial modification of adjectives. Our adjective-noun phrases contained both intersective (e.g. 'round cookie', which is both round and a cookie) and non-intersective adjectives (a scalar adjective such as 'large' has a very different interpretation depending on whether it is applied to a typically small entity, such as 'mosquito', or a typically large one such as 'elephant'; see Bierwisch, 1987; Kamp & Partee, 1995). Within the adverb-verb subtype, all stimuli were temporal adverbs (e.g. 'always', 'never', 'frequently'), and, to avoid ambiguity with verbs in the present tense (e.g. 'dreams' could be a verb or a noun), all verbs were in the past tense (e.g. 'whistled').

For argument saturation, we included verbs, prepositions and determiners with their noun arguments. For simplicity, we constrained some of the properties of the items. For example, within the verb-noun subtype, while the verbs varied in whether they were optionally or obligatorily transitive, they were all in the simple present (e.g. 'repairs furniture'), in order to ensure that they were all understood as being in the active voice (as opposed to the past tense 'repaired furniture', which can be interpreted as a passive participle). Furthermore, so that we could use the two-word paradigm without the addition of determiners or plural suffixes, we used only mass nouns. In our preposition-noun subtype, we avoided the need for determiners and plural suffixes by using proper noun target words, such as place names (e.g. 'in Italy') and names of famous people (e.g. 'by Tolstoy'). Lastly, within the determiner-noun subtype, all determiners were possessive determiners and all nouns were singular.

We created 50 different target words for each syntactic subtype, for a total of 300 unique target words. Target words did not significantly differ in length (saturation M = 6, SD = 1.9, modification M = 6, SD = 1.8, $p > 0.5$). Log frequencies of the stimuli were determined using the Google Web 1T 5-gram corpus (Brants & Franz, 2006). Although we were not interested in any main effects of target word type, we do note that argument saturation target words (M = 7.4, SD = 0.8) were slightly but significantly more frequent than modification target words (M = 7.1, SD = 0.8, $p < 0.01$). However, since we were comparing activity elicited by the same target word in combinatory and non-combinatory contexts, any frequency differences between target words would be expected to affect both combinatory and non-combinatory conditions equally and should not, therefore, affect the combinatory processes per se. Crucially, the transition probability between the first word and the target word was matched between the combinatory types (saturation M = 0.003, SD = 0.009, modification M = 0.006, SD = 0.03, $p = 0.25$).

Target words were presented once each within their compositional phrase (two-word combinatory condition, e.g. 'black sweater') and alone (one-word non-combinatory condition, e.g. 'rkgjg sweater'). In order to match the amount of visual stimulation as closely as possible in both conditions, and to avoid any effect of surprise or novelty in response to the presentation of a word in isolation, words in the non-combinatory condition were preceded by

Subcondition	Combinatory	Non-Combinatory
Modification		
Adjective-Noun	*black sweater*	*rkgjg sweater*
Adverb-Verb	*never jogged*	*nhcny jogged*
Adverb-Adjective	*very soft*	*rmwz soft*
Argument Saturation		
Verb-Noun	*eats meat*	*trwq meat*
Preposition-Noun	*in Italy*	*xq Italy*
Determiner-Noun	*Tarzan's vine*	*fkbczswh vine*

Fig. 1. Experimental design for English.

unpronounceable consonant strings matched in length to the first word in the two-word condition (Bemis & Pylkkänen, 2011). Though this design involved repetition of the target items, stimuli were presented in a random order to each participant. Therefore, if there were effects of repetition (e.g. Stowe et al., 1998), we would not expect them to lead to differences between experimental conditions.

2.1.2. Arabic

The Arabic experiment was designed as the English experiment except in certain small respects, which we highlight here. In Arabic, we kept the paradigm simple by using only verb-noun argument saturation and adjective-noun modification. Because adjectival modification in Arabic is typically post-nominal (‘boat red’ instead of ‘red boat’), Arabic allowed us to test the robustness of combinatory effects to word order variation (in all cases of modification in the English design, the modifier preceded the head of the phrase, Fig. 1). However, in some instances adjectival modifiers do precede the noun in Arabic, such as in superlative modification (e.g. ‘biggest ship’), and thus we were also able to include an adjective-noun condition that matched the English word order. In total, we had three subconditions: verb-noun argument saturation, noun-adjective modification, and superlative-noun modification (see Fig. 2 for design and stimulus examples. Note that Arabic is written from right to left, so the word or nonword on the right was displayed first). All Arabic verbs were in the present tense (e.g. يقفز ‘jumps’) in order to avoid ambiguity with verbs in the past tense as those could be read as a verb or a noun (e.g. ركض could be read as the noun رَكْض ‘the act of running’ or as the verb رَكَضَ ‘ran’). Just as in English, we included 50 unique phrases in each subcondition, for a total of 150 unique target words. Transition probabilities were matched across rule types (saturation M = 0.01, SD = 0.03, modification M = 0.003, SD = 0.01, $p = 0.16$) using frequencies from the Arabic Gigaword corpus (Graff, 2007). All items were in Modern Standard Arabic.

In Arabic, when the target words were presented alone they were always preceded by an unpronounceable consonant string. The strings, which were phonologically illicit nonwords, were made up of individual unconnected characters to discourage participants from attempting to interpret them as words.

2.2. Participants

2.2.1. English

29 healthy, right-handed, native English speakers participated in the first experiment (10 male; M = 23 years, SD = 5) at the New York site of the NYU Neuroscience of Language Lab. All had normal or corrected-to-normal vision and gave informed consent to participate. Five subjects were excluded from further analysis for performing below 75% accuracy, and a further six were excluded because of excess noise in their raw data (defined as over 40% loss of trials in at least one condition after artifact rejection). 18 subjects were included in the final stages of analysis.

2.2.2. Arabic

27 healthy, right-handed, native Arabic speakers participated in the second experiment (17 male; M = 21 years, SD = 1) at the Abu Dhabi site of the NYU Neuroscience of Language Lab. Participants came from a variety of linguistic and educational backgrounds, though all had native competence in Modern Standard Arabic. All had normal or corrected-to-normal vision and gave informed consent to participate. One subject was excluded from further analysis for performing at 50% accuracy, and a further three were excluded because of excess noise in their raw data (defined as over 40% loss of trials in at least one condition after artifact rejection). 23 subjects were included in the final stages of analysis.

2.3. Task

In both experiments, stimuli were presented in pseudorandom order and all subjects saw all stimuli. In 20% of trials, evenly distributed across subtypes and conditions, a comprehension task was included to ensure that the subjects were paying attention and participating actively. At the end of these task trials, a short phrase appeared on the screen and remained onscreen until subjects had indicated whether the phrase was a “match” or a “mismatch” to the preceding stimulus trial, by evaluating whether it made sense in the context of the trial. For example, after the trial stimulus ‘kjxqk knitted’, the comprehension phrase was ‘scarves’, at which point the subject was expected to respond ‘yes’. 50% of the task items made sense in the context, and 50% did not. Note that most task questions did not have a perfectly unambiguous right or wrong answer; rather, they depended on the participant’s subjective evaluation of whether a question trial such as ‘scarves’ made sense in the context of the word ‘knitted’. Accuracy was coded as agreement with the experimenter. Because of this, the accuracy limit for inclusion in analysis was set rather low, at 75% for English and 50% for Arabic (our Arabic participants were less used to participating in experiments so we set a lower threshold for inclusion). Due to the simple attention-monitoring nature of the task, behavioral responses were not analyzed further, as they were not expected to reflect experiment-relevant processes.

2.4. Procedure

2.4.1. English

Prior to recording, the subjects’ head shapes were digitized using a Polhemus Fastrak three-dimensional digitizer. Marker coils located at five positions around the face were digitized, and their position was measured with respect to the MEG sensors. This allowed us to determine position of the subject inside the dewar, as well as constrain the source localization by coregistering

Subcondition	Combinatory	Gloss	Non-Combinatory	Gloss
Noun-Adjective	قطار سريع	*train fast*	ك ج خ سريع	*xbql fast*
Adjective-Noun	أكبر سفينة	*biggest ship*	ه ب س ص سفينة	*nmkfs ship*
Verb-Noun	يأكل لحم	*eats meat*	أ ج خ لحم	*kdqm meat*

Fig. 2. Experimental design for Arabic.

individual head shapes to a standard smoothed brain in BESA 5.1. (MEGIS Software GmbH, Gräfelfing, Germany). Furthermore, subjects were fitted with EOGs that recorded eye movements, which were then used for blink artifact rejection. Participants practiced the task on a shortened block containing stimuli distinct from the experimental materials before the MEG recording.

Subjects lay in a dimly-lit, magnetically-shielded room and performed the task in five separate blocks (stimuli were randomized separately for each subject, and blocks were presented in random order). MEG data were collected using a 157-channel whole-head axial gradiometer system (Kanazawa Institute of Technology, Nonoichi, Japan) sampled at 1000 Hz with a low-pass filter at 200 Hz and a notch filter at 60 Hz. Stimuli were presented using PsychToolBox software (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and projected onto a screen approximately 50 cm away. Words were presented for 300 ms each, in white 30-point Courier font, on a grey background. A blank screen was presented for 300 ms between words. Task words remained on the screen until subjects indicated their response (Match or No Match) by pressing a button with either the index or middle finger of their left hand. The inter-trial interval was normally distributed, with a mean of 400 ms (SD = 100 ms). The recording session, including preparation time and practice outside the scanner, lasted approximately an hour and a half.

2.4.2. Arabic

Procedure was in all important respects identical to the English experiment, except that the experiment was divided into four blocks rather than five, and the Arabic lab has a 207-channel whole-head axial gradiometer system (Kanazawa Institute of Technology, Nonoichi, Japan). Also, words were saved as images prior to presentation, using a white 48-point Nazli font.

2.5. Analysis

2.5.1. English

Raw data were highpass filtered at 0.1 Hz prior to epoching. Data for each subject were segmented into epochs with a 200 ms pre-stimulus interval and a 700 ms post-stimulus interval. Data were cleaned of artifacts by rejecting trials for which the maximum amplitude exceeded 3000 fT or in which the subjects blinked (as determined manually). An activity baseline (channel noise covariance matrix) was taken from the 100 ms before target word onset. We performed a region of interest (ROI) analysis on a region covering most of the left temporal lobe (LTL), which was followed by a whole-brain source analysis primarily aimed at verifying that effects seen in the ROI analyses in fact reflected activity in the targeted regions as opposed to spillover from adjacent areas.

2.5.1.1. Minimum norm estimates. We constructed distributed L2 minimum norm source estimates for each subject and for each condition in BESA 5.1 (MEGIS Software GmbH, Gräfelfing, Germany). This method places two shells, each containing 713 evenly distributed regional sources, at 10% and 30% below a standard smoothed brain surface. Each regional source contains two orthogonally-oriented dipoles, and the total activity of the source is the root mean square (RMS) of the two dipoles. After this RMS, the regional source with the largest value is selected at each location. Minimum norm images were depth and spatiotemporally weighted, using a signal subspace correlation measure (Mosher & Leahy, 1998). After artifact rejection, the individual subconditions ended up being too underpowered for reliable source estimation (often 30 or fewer trials per condition) and consequently, we decided to maximize our power by instead combining the epochs of all three subconditions of each rule type into a single condition, yielding a 2 × 2 design with rule type (modification vs. saturation) and number of words (one vs. two) as factors.

2.5.1.2. ROI analysis. Though MEG has high temporal resolution, its spatial resolution is lower than that of PET or fMRI. Therefore, we included a relatively large region of cortex in our ROI. We did this both to compensate for the lower spatial resolution of MEG and also because the sentence processing and semantic memory literatures have not yet established clear anatomical guidelines for what constitutes the LATL; in fact, the term LATL often refers to the location of damage in semantic dementia patients (e.g. Patterson, Nestor, & Rogers, 2007) despite the fact that this consists of a relatively broad region that almost certainly contains several functional subdivisions. Therefore, we cast a wide spatial net and constructed an ROI that covered most the left lateral temporal lobe (LTL). We achieved this by assigning Brodmann area labels to the 713 sources on the smooth BESA cortex using the Tailarach daemon (Lancaster et al., 1997, 2000) and including Brodmann areas 38, 20, and 21 in our ROI. We then ran 2 × 2 cluster permutation ANOVAs (Maris & Oostenveld, 2007) with rule type (modification and argument saturation) and number of words (two words or one word) as factors. We then extracted activity over any identified clusters for every subcondition in order to investigate whether individual subconditions were driving the identified effects. For our cluster permutation tests, we followed Bemis and Pylkkänen (2011) in only including clusters that maintained a significance of at least $p = 0.3$ for 10 or more consecutive milliseconds. We constrained our time window of interest to 0–600 ms after presentation of the noun, and used the F values from the repeated measures ANOVA as our statistic of interest.

2.5.1.3. Whole brain analysis. In order to confirm that our ROI analysis was correctly localizing combinatory responses rather than capturing spillover from adjacent regions, we performed whole brain comparisons between the two and one word conditions within each rule type. For each source, we performed a paired t-test at each time point in each rule type. We plotted effects that were significant at $p < 0.05$, uncorrected for multiple comparisons, had at least 5 significant neighboring sources, and maintained significance for at least 5 consecutive time points, on the smoothed BESA brain. This test is used only as a confirmation of our ROI results, as it is too liberal to allow us to draw independent conclusions.

2.5.1.4. Cross-composition-type validation analysis. Our large ROI did not allow us to draw the conclusion that we are capturing effects in identical regions across rule types. Therefore, we performed a cross-composition-type validation analysis in order to determine whether the sources that show combinatory effects in one rule type also capture combinatory effects in the other rule types. This was accomplished by constructing separate functional ROIs based on each composition type, and testing combinatory responses in these ROIs for the other composition types. This analysis proceeded as follows: for each composition type within a language, we constructed an ROI that captured the center of combinatorial activity in that composition type, which was defined as the sources in the left temporal lobe that survived the significance criteria of the uncorrected whole brain comparison (as described above in Section 2.5.1.3) for the interval of 200–300 ms. We then conducted one-tailed cluster permutation *t*-tests in this ROI between the two-word and one-word conditions of all other composition types within the language, correcting for multiple comparisons between 100 and 400 ms. This cross-validation provided a way to investigate the spatial overlap between composition responses in different composition types by independently constructing an ROI from one condition and running corrected permutation tests in the other conditions.

2.5.2. Arabic

Analysis procedures were identical except that Arabic data were not highpass filtered as the recording environment was less noisy, and given that we had three combinatory contexts, we ran 3×2 cluster permutation ANOVAs with combinatory context (argument saturation, adjective-noun modification, noun-adjective modification) and number of words (two words or one word) as factors.

2.5.3. Cross-language analysis

We were also interested in investigating whether there was spatial overlap between combinatorial responses cross-linguistically. Therefore, we performed a cross-language analysis following the same methods as for our cross-composition-type analysis (Section 2.5.1.4). Generally, this analysis consisted of testing combinatory responses in one language using ROIs defined on the basis of the other language. For both English and Arabic, we combined all left temporal sources between 200 and 300 ms across all composition types in one language (again using the criteria from our whole-brain analysis, Section 2.5.1.3) into a single ROI, and then performed one-tailed cluster permutation *t*-tests within this ROI across all composition types in the second language, correcting for multiple comparisons between 100 and 400 ms.

3. Results

3.1. Behavioral results

Our attention-monitoring sensicality judgment task yielded an overall accuracy of 86.7% (SD = 5.5%) across 18 English-speaking subjects and 88% (SD = 8%) across 24 Arabic-speaking subjects. Because our task was designed strictly as a way to engage participants' attention, accuracy did not directly assess the behavioral cost of composing the presented words. Therefore, we did not consider the behavioral responses further.

3.2. ROI results

3.2.1. English

Cluster permutation tests in our LTL ROI yielded an extended cluster showing a main effect of number of words that extended from 35 ms to 453 ms ($p = 0.0049$, see Fig. 3), and no significant clusters for an interaction between rule type and number of words or for a main effect of rule type. Follow-up two-tailed permutation *t*-tests in each rule type showed a long cluster of increased activity in the two-word condition for the argument saturation rule type (61–340 ms, $p = 0.0134$), and a shorter, later cluster of combinatory activity in the modification rule type which was, however, not significant after correction for multiple comparisons over the interval of 0–600 ms (205–300 ms, $p = 0.4$).

3.2.2. Arabic

In the Arabic experiment, cluster permutation tests yielded a cluster extending from 156 ms to 348 ms that showed a main effect of number of words ($p = 0.025$, see Fig. 4), and no significant clusters for an interaction between number of words and rule type. Follow-up two-tailed *t*-test cluster permutation tests showed a cluster of significantly increased combinatorial activity within the main effect time window for the noun-adjective modification condition (139–312 ms, $p = 0.046$). Both other composition types also showed a cluster within the same time window, though the clusters were not significant after cluster permutation correction over 0–600 ms (argument saturation: 196–300 ms, $p = 0.2$; superlative-noun modification: 209–271 ms, $p = 0.4$). In Arabic as in English, the qualitative trend regarding timing was that the peak of the combinatory response was earliest in the argument saturation condition.

3.3. Whole brain comparisons

3.3.1. English

See Fig. 5 for whole brain comparisons between the two word and one word conditions for both argument saturation and modification. In the argument saturation condition, there was a strong early combinatory response localized in the LTL, and little evidence for a combinatory response in other regions. In the modification condition there was a later cluster of sources showing a combina-

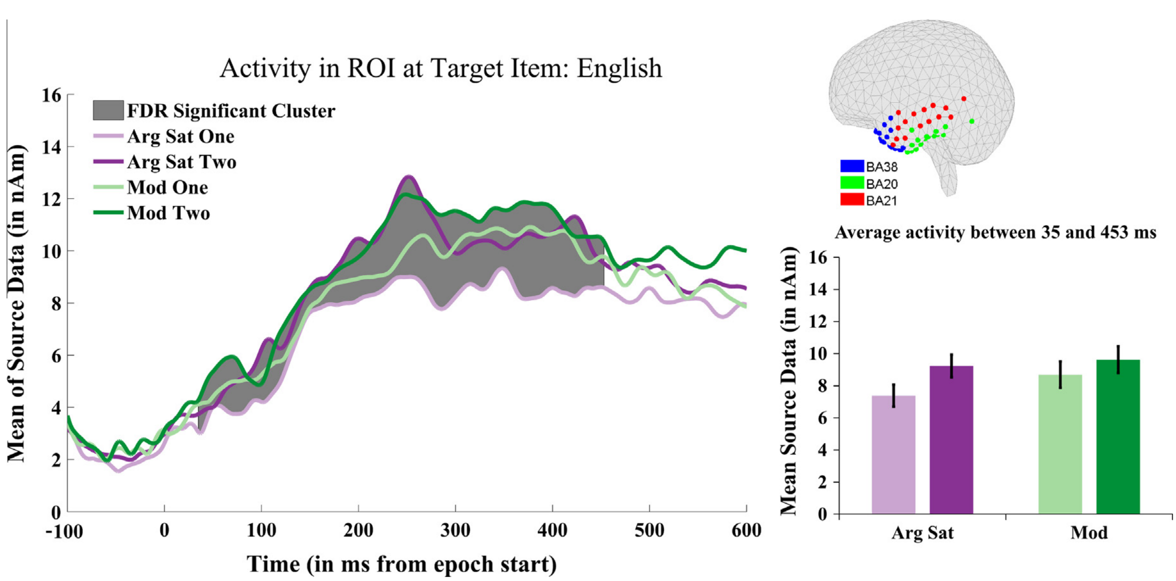

Fig. 3. Activation (in nAm) by condition in the LATL ROI for the English experiment. 0 ms represents onset of the target word. A main effect of number of words was found between 35 and 453 ms. Chart to the right shows average activity over the time window for each condition. Error bars show SEMs.

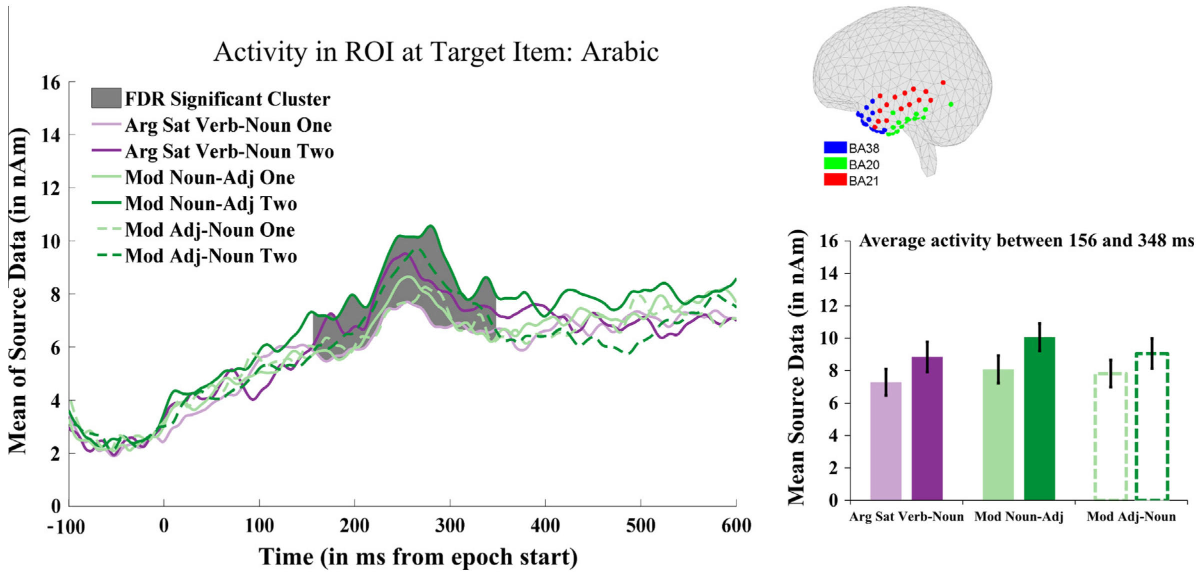

Fig. 4. Activation (in nAm) by condition in the LATL ROI for the Arabic experiment. 0 ms represents onset of the target word. A main effect of number of words was found between 156 and 348 ms. Chart to the right shows average activity over the time window for each subcondition. Error bars show SEMs.

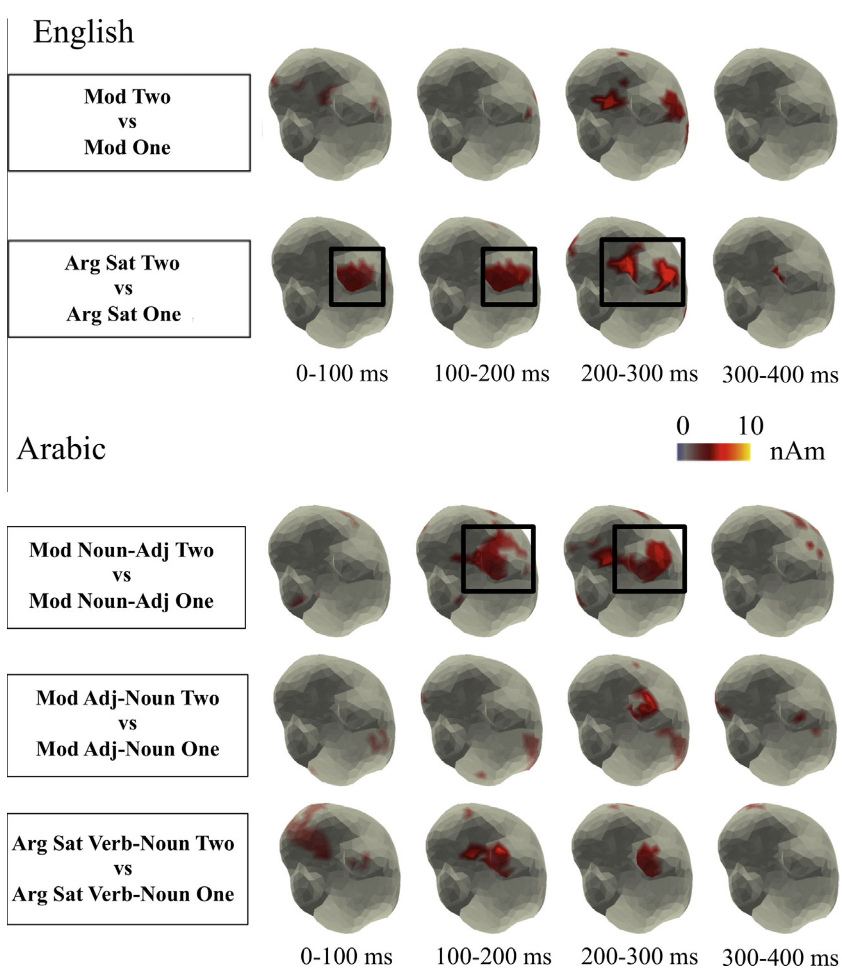

Fig. 5. Top: Whole brain comparisons of activity after onset of the target noun for the English experiment. Activity is shown (in red) if the first condition elicited greater activity than the second condition in at least 5 adjacent sources, and lasted for a minimum of 5 ms, at $p < 0.05$ (uncorrected). Bottom: Whole brain comparisons of activity after onset of the target noun for the Arabic experiment. Activity is shown (in red) if the first condition elicited greater activity than the second condition in 5 adjacent sources, and lasted for a minimum of 5 ms, at $p < 0.05$ (uncorrected).

tory response pattern in the LTL, though recall that the corresponding activity did not survive correction for multiple comparisons in the ROI analysis. Generally, these spatial clusters were more posterior in English than in Arabic, particularly in the modification condition.

3.3.2. Arabic

The whole brain comparisons between the two-word and one-word conditions in every subcondition (see Fig. 5) showed clusters of sources exhibiting a combinatory response pattern in the LTL, though recall that in the ROI analysis only the noun-adjective modification condition showed a significant effect. No other regions showed a consistent combinatory response. Overall, our whole brain comparisons in both experiments supported the conclusions of our ROI analysis and highlighted the fact that combinatory responses were primarily localized in the LATL.

3.4. Cross-composition-type validation analysis

3.4.1. English

In the cross-composition-type analysis, ROIs were defined on the basis of effects observed in one composition type to examine whether effects for the other composition type would be observed in exactly the same sources. 11 sources were included in the ROI for the modification composition type (see Fig. 6 for sources and waveforms in all conditions). One-tailed cluster permutation *t*-tests in this ROI showed a cluster of significant combinatorial activity in the argument saturation condition (134–301, $p = 0.007$). For argument saturation, the whole brain comparison indicated two spatially non-adjacent clusters within the left anterior temporal lobe: one lateral cluster containing 12 sources and a separate ventral cluster containing 3 sources. We included these two clusters as two separate ROIs in the analysis. For the lateral ROI, there was a significant combinatory response in the modification rule type: 150–291 ms, $p = 0.04$. The ventral ROI, however, showed no significant clusters in the modification condition (110–137 ms, $p = 0.4$, 357–400 ms, $p = 0.33$). Thus, on the lateral surface of left anterior temporal cortex, we observed areas of overlap between significant combinatory responses across composition type, and on the ventral surface, a cluster of sources that mostly only responded to argument saturation. As reported below, however, this potential argument saturation-specific finding did not generalize to Arabic.

3.4.2. Arabic

27 sources were included as an ROI based on effects observed for the noun-adjective condition. Cluster permutation tests within this ROI showed significant clusters of increased combinatorial activity in the superlative-noun modification condition

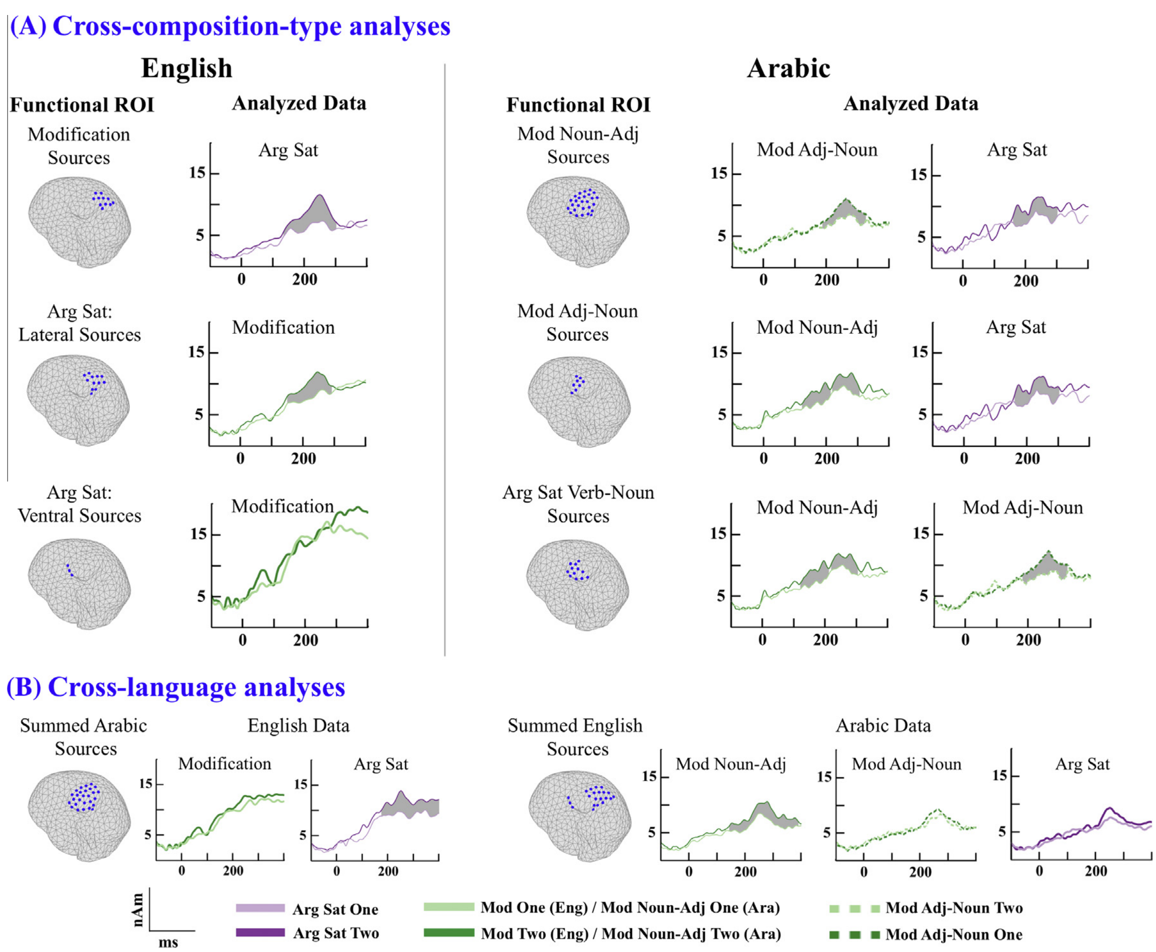

Fig. 6. Sources and waveforms within all functional ROIs for the cross-composition-type (A) and cross-language analysis (B). Sources included in each ROI, as determined using selection criteria from the whole brain analysis, are shown for each comparison in the leftmost column of each language. Waveforms are organized in columns by language and composition type and in rows by ROI, and show activation (in nAm) by condition in the ROI. Shaded regions represent significant increases in the two-word condition as compared to a one-word control, as determined by one-tailed cluster permutation tests. Significant combinatory responses were seen across all composition types for each composition type ROI in Arabic, and in all but the ventral argument saturation ROI in English. Across languages, the ROI based on summed Arabic sources showed a significant combinatory effect only in the English argument saturation condition, and the ROI based on the summed English sources showed a significant combinatory effect only in the Arabic noun-adjective modification condition.

(189–325 ms, $p = 0.02$) and the argument saturation condition (161–303 ms, $p = 0.03$). The same was true for 8 sources that were identified as the superlative-noun modification ROI (noun-adjective: 123–213 ms, $p = 0.03$, 219–305 ms, $p = 0.04$; argument saturation: 162–303 ms, $p = 0.008$) and for 12 sources that constituted the argument saturation ROI (superlative-noun modification: 183–325, $p = 0.008$; adjective-noun modification, 122–212 ms, $p = 0.02$, 220–304 ms, $p = 0.04$). In sum, this analysis showed clear spatial overlap between combinatory responses across all rule types. No argument saturation-specific activity was observed in the Arabic data.

3.5. Cross-language analysis

For the English-based ROI, we constructed a region containing the sum of all sources showing either an argument saturation or a modification effect in English, for a total of 21 sources. Cluster permutation tests in this ROI across the three Arabic subconditions showed clusters of increased activity for the combinatory condition in each subcondition; however, this effect was significant after correction for multiple comparisons only for noun-adjective modification (144–389 ms, $p = 0.007$; superlative-noun modification: 215–232 ms, $p = 0.5$, 250–264 ms, $p = 0.6$, 305–324 ms, $p = 0.51$; argument saturation: 213–266 ms, $p = 0.19$).

The similarly constructed summed Arabic ROI contained 27 sources. We found a significant combinatory effect in this ROI for the English argument saturation condition (175–400 ms, $p = 0.008$), and non-significant clusters in the English modification condition (117–141 ms, $p = 0.38$, 218–249 ms, $p = 0.36$). Generally, because the center of the Arabic combinatory response was localized more anteriorly than in English, the summed Arabic sources best captured the English argument saturation effect as it included the small ventrally localized effect. In contrast, the summed English sources were too posterior to robustly capture Arabic effects other than noun-adjective modification, which appeared to spread the most posteriorly (see whole brain comparisons in Fig. 5).

3.6. Results summary

Overall, our results rule out the hypothesis that the LATL effects observed in prior studies are computationally specific to modification type environments alone. In fact, though we found a main effect of number of words, and no interaction between composition types, in English, only argument saturation elicited significant LTL responses in follow-up t-tests. However, results from our cross-composition-type validation analysis showed significant (corrected) composition effects for each composition type in smaller, independently selected regions contained within our larger LTL ROI (see Section 4). Taken together, these results constitute stronger evidence for similarity as opposed to difference between the modification and saturation rule types.

In Arabic, we found a main effect of number of words, and no interaction between composition types. In follow-up *t*-tests, though only noun-adjective modification elicited significant LTL effects in combinatory environments, significant corrected effects were seen in each composition type in smaller regions within the larger ROI. As in English, we believe that these results point to convergence across rule types.

In sum, in both languages evidence was obtained that both modification and argument saturation increase LATL activity in combinatory contexts. A cross-composition-type analysis confirmed that similar sources were driving combinatory activity in each composition type, and this was true cross-linguistically. Another descriptive commonality to the two datasets was that the onset of composition effects was earlier for argument saturation than modification in both languages. One possible difference between the languages was that the center of combinatory activity appeared to be more anterior in Arabic than in English.

4. Discussion

In this pair of experiments, we set out to expand current knowledge of the combinatory profile of the LATL by investigating responses to phrases composed by both modification and argument saturation across two languages. In the ROI analyses of both experiments, we found main effects of the number of words presented, with two-word conditions that involved composed phrases eliciting more activity than single words, and no interactions between rule type and number of words. In follow-up t-tests in English, the only condition showing a significant combinatory effect in our larger ROI was argument saturation. In Arabic, the only condition showing a significant combinatory effect in our larger ROI was noun-adjective modification. However, our LTL ROI was quite large, including all sources in Brodmann areas 20, 21, and 38. We constructed such an inclusive ROI because we did not have a strong a priori spatial hypothesis about the localization of combinatory effects. Yet in so doing we undoubtedly included many non-combinatory sources, thereby diminishing our power to detect significant combinatory effects. This may well have contributed to the lack of combinatory effects in each subcondition. Support for this reasoning comes from the fact that, when we constructed smaller functional ROIs within this larger ROI based on independent contrasts in other rule types, we found significant combinatory effects in every rule type in both languages. This analysis points to smaller clusters within our larger ROI that show overlapping combinatory effects across rule types.

These results substantially expand the combinatory profile of the LATL, pointing to a unified combinatory effect across conditions, despite rule type and word order differences. Generally, combinatory responses were quite similar cross-linguistically. It is worth noting that the combinatory responses in the English experiment had a very rapid onset and were long lasting. Though the beginning of the cluster of significant activity may have started earlier than previous experiments, importantly, the peak of activity in both rule types was at around 250 ms (earlier in the argument saturation condition), which is consistent with previous work using this paradigm (Bemis & Pylkkänen, 2011, 2013). The location of peak activity revealed another slight difference between English and Arabic. Whole brain analyses showed that English combinatory responses were located more posteriorly than in Arabic, and the cross-language analysis confirmed this. Though we may not have strong evidence that combinatory responses localize to the exact same regions cross-linguistically, we still believe that this is the most plausible interpretation of these effects. Despite showing slight localization differences, the composition responses were quite similar in wave morphology and peak timing, which suggests a similar process occurring across languages. Whether or not we are capturing identical regions across languages, in light of these results, it is highly likely that previous experiments showing engagement of the LATL by structured, well-formed sentences were in fact capturing basic combinatory activity rather than more complex sentence-level operations.

Though we found no compelling evidence supporting a distinction in the LATL between modification and argument saturation-type composition, we did find slight latency differences between the two composition types in both languages. Though the differences are descriptive only, we observed that in both languages, the peak of the combinatory response appeared to be slightly earlier in the argument saturation conditions than for the modification conditions. The psycholinguistic literature on arguments and adjuncts suggests that there may be a slight advantage in the processing of arguments (see, for example, Boland & Boehm-Jernigan, 1998; Clifton, Speer, & Abney, 1991; Kennison, 2002; Liversedge, Pickering, Branigan, & van Gompel, 1998; Schütze & Gibson, 1999; Speer & Clifton, 1998), possibly due to a parser bias towards the former (Abney, 1989; Pylkkänen & McElree, 2006; Schütze & Gibson, 1999). Our results appear, at least descriptively, to be compatible with this interpretation of the argument advantage.

Finally, much of the sentence-level literature that has highlighted the role of the LATL in composition has in fact implicated a network of regions in semantic processes (Binder, Desai, Graves, & Conant, 2009). This network typically include the angular gyrus (Bavelier et al., 1997; Bottini et al., 1994; Humphries et al., 2006, 2007), the left inferior frontal gyrus (Baggio & Hagoort, 2011; Hagoort, 2005; Hagoort, Baggio, & Willems, 2009; Humphries et al., 2006, 2007), or the ventromedial prefrontal cortex (Bemis & Pylkkänen, 2011; Pylkkänen, 2008; Pylkkänen, Martin, McElree, & Smart, 2009; Pylkkänen & McElree, 2007; Pylkkänen, Oliveri, & Smart, 2009). Intriguingly, when using a more targeted paradigm to investigate basic composition, we found no evidence for a consistent involvement of these regions in general combinatorics, across composition types and languages. Though we cannot rule out the involvement of other regions in basic composition, these results suggest that the LATL is the best candidate for the primary locus of a basic conceptual combinatory mechanism.

In conclusion, by investigating two distinct composition types, modification and argument saturation, in two languages, Arabic and English, we demonstrated that the LATL is sensitive to composition in a variety of constructions in two very different languages. The generality of the response suggests that the LATL is a central site for composing words or concepts regardless of their semantic or syntactic relationship.

Acknowledgments

This research was supported by the National Science and Engineering Research Council of Canada (M.W.), the Fonds de recherche du Québec-Nature et technologies (M.W.), the National Science Foundation Grants BCS-0545186 and BCS-1221723 (L.P.), grant G1001 from the NYUAD Institute, New York University Abu Dhabi (L.P.), and the Whitehead Fellowship for Junior Faculty Biomedical and Biological Sciences (L.P.)

References

Abney, S. P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research, 18*(1), 129–144.
Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes, 26*(9), 1338–1367.
Baron, S. G., & Osherson, D. (2011). Evidence for conceptual combination in the left anterior temporal lobe. *NeuroImage*.
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.
Bavelier, D., Corina, D., Jezzard, P., Padmanabhan, S., Clark, V. P., Karni, A., et al. (1997). Sentence reading: A functional MRI study at 4 Tesla. *Journal of Cognitive Neuroscience, 9*(5), 664–686. http://dx.doi.org/10.1162/jocn.1997.9.5.664.
Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *The Journal of Neuroscience, 31*(8), 2801.
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*.
Bierwisch, M. (1987). The semantics of gradation. In M. Bierwisch & E. Lang (Eds.), *Dimensional adjectives*. Berlin: Springer.
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. (1460–2199 (Electronic)).
Boland, J. E., & Boehm-Jernigan, H. (1998). Lexical constraints and prepositional phrase attachment. *Journal of Memory and Language, 39*(4), 684–719.
Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., et al. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language: A positron emission tomography activation study. *Brain, 117*(6), 1241.
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.
Brants, T., & Franz, A. (2006). Google Web 1T 5-gram Version 1. In L. D. Consortium (Ed.). Philadelphia.
Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkänen, L. (2010). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language, 120*(2), 163–173.
Clifton, C., Jr, Speer, S., & Abney, S. P. (1991). Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language, 30*(2), 251–271.
Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science, 24*(2), 299–349.
Del Prato, P., & Pylkkanen, L. (2014). MEG evidence for conceptual combination but not numeral quantification in the left anterior temporal lobe during language production. *Language Sciences, 5*, 524.
Dronkers, N. F., & Wilkins, D. P. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition, 92*(1), 145–177.
Friederici, A. D., Meyer, M., & Von Cramon, D. Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language, 74*(2), 289–300.
Gorno-Tempini, M. L., Rankin, K. P., Woolley, J. D., Rosen, H. J., Phengrasamy, L., & Miller, B. L. (2004). Cognitive and behavioral profile in a case of right anterior temporal lobe neurodegeneration. *Cortex, 40*(4–5), 631–644.
Graff, D. (2007). Arabic Gigaword Third Edition. In L. D. Consortium (Ed.). Philadelphia.
Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9*(9), 416–423.
Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. *The Cognitive Neurosciences*, 819–836.
Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., et al. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage, 17*(3), 1101–1116.
Hampton, J. A. (1997). Conceptual combination. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 133–159). Cambridge, MA: MIT Press.
Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell.
Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Progressive fluent aphasia with temporal lobe atrophy. *Brain, 115*(6), 1783–1806.
Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience, 18*(4), 665–679.
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. http://dx.doi.org/10.1016/j.neuroimage.2007.03.059.
Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Human Brain Mapping, 26*(2), 128–138.
Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition, 57*(2), 129–191. http://dx.doi.org/10.1016/0010-0277(94)00659-9.
Kennison, S. M. (2002). Comprehending noun phrase arguments and adjuncts. *Journal of Psycholinguistic Research, 31*(1), 65–81.
Kho, K. H., Indefrey, P., Hagoort, P., Van Veelen, C., Van Rijen, P., & Ramsey, N. (2008). Unimpaired sentence comprehension after anterior temporal cortex resection. *Neuropsychologia, 46*(4), 1170–1178.
Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*, 14.
Lancaster, J., Rainey, L., Summerlin, J., Freitas, C., Fox, P., Evans, A., et al. (1997). Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method. *Human Brain Mapping, 5*(4), 238.
Lancaster, J., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping, 10*(3), 120–131.
Lau, E. F., Gramfort, A., Hämäläinen, M. S., & Kuperberg, G. R. (2013). Automatic semantic facilitation in anterior temporal cortex revealed through multimodal neuroimaging. *The Journal of Neuroscience, 33*(43), 17174–17181.
Liversedge, S. P., Pickering, M. J., Branigan, H. P., & van Gompel, R. P. (1998). Processing arguments and adjuncts in isolation and context: The case of by-phrase ambiguities in passives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(2), 461.
Magnusdottir, S., Fillmore, P., den Ouden, D., Hjaltason, H., Rorden, C., Kjartansson, O., et al. (2013). Damage to left anterior temporal cortex predicts impairment of complex syntactic processing: A lesion-symptom mapping study. *Human Brain Mapping, 34*(10), 2715–2723.
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190.
Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., et al. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience, 5*(4), 467–479.
Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology, 20*(2), 158–190.

Montague, R. (1974). Formal philosophy; selected papers of Richard Montague.
Mosher, J. C., & Leahy, R. M. (1998). Recursive MUSIC: a framework for EEG and MEG source localization. *IEEE Transactions on Biomedical Engineering, 45*, 1342–1354.
Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language, 29*(3), 259–288.
Nobre, A. C., & Mccarthy, G. (1995). Language-related field potentials in the anterior-medial temporal lobe: II. Effects of word type and semantic priming. *The Journal of Neuroscience, 15*(2), 1090–1098.
Noppeney, U., Price, C. J., Duncan, J. S., & Koepp, M. J. (2005). Reading skills after left anterior temporal lobe resection: An fMRI study. *Brain, 128*(6), 1377–1385.
Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition, 9*(1), 35–58. http://dx.doi.org/10.1016/0010-0277(81)90013-5.
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.
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.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.
Pylkkänen, L. (2008). Mismatching meanings in brain and behavior. *Language and Linguistics Compass, 2*(4), 712–738.
Pylkkänen, L., Bemis, D. K., & Blanco Elorrieta, E. (2014). Building phrases in language production: An MEG study of simple composition. *Cognition, 133*(2), 371–384.
Pylkkänen, L., Brennan, J., & Bemis, D. K. (2011). Grounding the cognitive neuroscience of semantics in linguistic theory. *Language and Cognitive Processes, 26*(9), 1317–1337. http://dx.doi.org/10.1080/01690965.2010.527490.
Pylkkänen, L., Martin, A. E., McElree, B., & Smart, A. (2009). The anterior midline field: Coercion or decision making? *Brain and Language, 108*(3), 184–190.
Pylkkänen, L., & McElree, B. (2006). The syntax-semantics interface: On-line composition of sentence meaning. *Handbook of Psycholinguistics*, 537–577.
Pylkkänen, L., & McElree, B. (2007). An MEG study of silent meaning. *Journal of Cognitive Neuroscience, 19*(11), 1905–1921.
Pylkkänen, L., Oliveri, B., & Smart, A. J. (2009). Semantics vs. world knowledge in prefrontal cortex. *Language and Cognitive Processes, 24*(9), 22.
Rogalsky, C., & Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex, 19*(4), 786.
Schütze, C. T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language, 40*(3), 409–431.
Smith, E. E., & Osherson, D. N. (1984). Conceptual combination with prototype concepts. *Cognitive Science, 8*(4), 337–361. http://dx.doi.org/10.1016/s0364-0213(84)80006-3.
Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science, 12*(4), 485–527. http://dx.doi.org/10.1207/s15516709cog1204_1.
Speer, S. R., & Clifton, C. (1998). Plausibility and argument structure in sentence comprehension. *Memory & Cognition, 26*(5), 965–978.
Stowe, L. A., Broere, C. A. J., Paans, A. M. J., Wijers, A. A., Mulder, G., Vaalburg, W., et al. (1998). Localizing components of a complex task: Sentence processing and working memory. *NeuroReport, 9*(13), 2995.
Westerlund, M., & Pylkkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia, 57*, 59–70.
Wilson, S. M., Galantucci, S., Tartaglia, M. C., & Gorno-Tempini, M. L. (2012). The neural basis of syntactic deficits in primary progressive aphasia. *Brain and Language, 122*(3), 190–198.
Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review, 4*(2), 167–183.
Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: Emergent features of word, sentence, and narrative comprehension. *NeuroImage, 25*(3), 1002–1015.