Simple Composition During Language Processing: 
An MEG Investigation

By

Douglas K. Bemis

A dissertation submitted in partial fulfillment 
of the requirements for the degree of 
Doctor of Philosophy 
Department of Psychology 
New York University 
September, 2012

_______________________________ 
Liina Pylkkänen – Advisor
Abstract

Decades of research in the cognitive neuroscience of language have produced a vast wealth of information about how the brain retrieves the meaning of single words or makes sense of complex sentences. The road in between these two processes, however, is distinctly less traveled. This dissertation represents an initial investigation into the combinatorial neural mechanisms that transform two simple words – an adjective (*red*) and a noun (*boat*), into a singular, complex mental representation (*red boat*). The first chapter lays the groundwork by using a minimal MEG paradigm to isolate neural activity related to basic combinatorial linguistic processing. The remaining chapters begin to explore the scope of the underlying combinatorial mechanisms in terms of modality, use, and domain. Though still far from bridging the gap from individual words to complete sentences, the following work presents a necessary first step towards reaching this goal.
# Table of Contents

Abstract................................................................................................................................. 1

Chapter 1: Simple Composition: A Magnetoencephalography Investigation into the
Comprehension of Minimal Linguistic Phrases (Published in the *Journal of Neuroscience*) 4

Abstract................................................................................................................................. 4
Introduction ............................................................................................................................ 6
Materials and Methods .......................................................................................................... 9
Results .................................................................................................................................. 26
Discussion .............................................................................................................................. 40
References .............................................................................................................................. 48
Figures .................................................................................................................................. 60

Chapter 2: Basic linguistic composition recruits the left anterior temporal lobe and left
angular gyrus during both listening and reading (Revision submitted to *Cerebral Cortex*) 69

Abstract................................................................................................................................. 69
Introduction ............................................................................................................................ 71
Materials and Methods .......................................................................................................... 84
Results .................................................................................................................................. 93
Discussion .............................................................................................................................. 102
References .............................................................................................................................. 113
Figures .................................................................................................................................. 126

Chapter 3: Simple composition during language production in MEG (Submitted to the
*Journal of Neuroscience*) ................................................................................................. 131

Abstract................................................................................................................................. 131
Introduction ............................................................................................................................ 133
Materials and Methods .......................................................................................................... 136
Chapter 4: Flexible Composition: MEG evidence for the deployment of basic combinatorial linguistic mechanisms in response to task demands

Abstract

Introduction

Methods and Materials

Results

Discussion

References

Figures
Chapter 1: Simple Composition: A Magnetoencephalography Investigation into the Comprehension of Minimal Linguistic Phrases (Published in the Journal of Neuroscience)

Abstract

The expressive power of language lies in its ability to construct an infinite array of ideas out of a finite set of pieces. Surprisingly, few neurolinguistic investigations probe the basic processes that constitute the foundation of this ability, choosing instead to focus on relatively complex combinatorial operations. Contrastingly, in the present work, we investigate the neural circuits underlying simple linguistic composition, such as required by the minimal phrase “red boat.” Using magnetoencephalography, we examined activity in humans generated at the visual presentation of target nouns, such as “boat,” and varied the combinatorial operations induced by its surrounding context. Nouns in minimal compositional contexts (“red boat”) were compared with those appearing in matched non-compositional contexts, such as after an unpronounceable consonant string (“xkq boat”) or within a list (“cup, boat”). Source analysis did not implicate traditional language areas (inferior frontal gyrus, posterior temporal regions) in such basic composition. Instead, we found increased combinatorial-related activity in the left anterior temporal lobe (LATL) and ventromedial prefrontal cortex (vmPFC). These regions have been linked previously to syntactic (LATL) and semantic (vmPFC) combinatorial processing in more complex
linguistic contexts. Thus, we suggest that these regions play a role in basic syntactic and semantic composition, respectively. Importantly, the temporal ordering of the effects, in which LATL activity (~225 ms) precedes vmPFC activity (~400 ms), is consistent with many processing models that posit syntactic composition before semantic composition during the construction of linguistic representations.
Introduction

Human language derives its unbounded expressive capacity from an ability to take simple, familiar building blocks (such as words) and combine them effortlessly into more complex representations, regardless of whether or not the result is familiar. Such productive composition is present in the comprehension of even minimal linguistic expressions. For example, the phrase blue flamingo conjures a coherent and meaningful mental representation despite a listener’s presumed unfamiliarity with such an object. Determining the neural bases of basic combinatorial operations, such as those underlying this simple construction, is therefore a central goal for the cognitive neuroscience of language and is a necessary precursor to understanding the comprehension of more complex expressions.

Interestingly, rather than focus on such elementary combinatorial operations, historically neurolinguistic studies have been driven primarily by questions about how people handle deviations from simple processing. Past experiments typically investigate complex syntactic structures, such as center-embedded or object-extracted relative clauses (Miller and Chomsky, 1963; Stromswold et al., 1996), or unexpected phrases, such as garden-path constructions (Bever, 1970; Osterhout et al., 1994) and implausible sentence completions (Kutas and Hillyard, 1980). In these manipulations, basic linguistic expressions serve as the ‘baseline,’ and effects are determined as differences in neural activity engendered by the more complex or surprising stimuli. Consequently, in these studies information about basic combinatorial operations can only be gained indirectly by arguing that effects observed during the comprehension
of more complex stimuli reflect differential processing by the same mechanisms responsible for comprehending basic constructions – a presupposition that may or may not be correct.

The few neurolinguistic investigations that reverse this norm and place basic combinatorial operations at the forefront have done so within the context of full sentences. Most often these studies compare the comprehension of sentences to that of unstructured lists of words or pseudowords (Friederici et al., 2000; Mazoyer et al., 1993). While effects elicited through such comparisons likely reflect basic combinatorial processes, it is unclear to what extent such results also include additional comprehension mechanisms, such as the establishment of reference and long-distance dependencies, the temporal anchoring of events, pragmatic inferences, the deployment of various types of memory, and potentially the construction of discourse representations (as most studies perform their analysis over multiple sentences at a time).

In the present work, we reduce the complexity of the critical stimuli to simple adjective-noun phrases (e.g. red boat) to investigate the neurophysiological foundations of elementary combinatoric processing. Using magnetoencephalography (MEG), we assessed increases in neural activity generated within minimal compositional contexts compared to matched non-compositional environments (Figure 1). The temporal resolution of MEG allowed us to examine combinatorial processing in specific, well-controlled linguistic contexts in which only the presence or absence of composition was varied. Provocatively, areas canonically implicated in traditional neurophysiological models of language processing appear to play a lesser role in basic
composition. In contrast, two areas recently discussed in numerous neurobiological studies, anterior temporal cortex and ventro-medial prefrontal cortex, are implicated in this critical and fundamental construction of linguistic structure and meaning.
Materials and Methods

Participants.

Twenty-five non-colorblind, native English speakers participated in the study (17 female, average age: 25.3). All had normal or corrected-to-normal vision and gave informed consent.

Experimental Design and Stimuli.

Our paradigm was designed to isolate neural activity associated with basic combinatorial mechanisms that operate during the comprehension of very simple linguistic phrases. We measured activity generated by the processing of common, object-denoting nouns presented either in a combinatorial phrasal context or in a non-combinatorial control context. Lexical operations elicited by the comprehension of multiple linguistic items (as in a compositional phrase) were controlled for by having participants perform two tasks, each consisting of a two-word and a one-word condition. The presence or absence of composition was then varied between tasks (Figure 1).

In the main “Composition” task, combinatorial contexts consisted of a color adjective followed by a simple noun (red boat), while in non-combinatorial contexts a length-matched, unpronounceable consonant string preceded the noun (xkq boat). To ensure attention to the critical words, participants were asked to indicate whether a following colored shape matched the preceding verbal material. To count as matching in the two-word trials (red boat), both the color and the shape of the picture had to
match the preceding words. In the one-word trials (xkq boat) only the shape of the picture was relevant for the decision.

As almost any meaningful word prior to the noun might elicit attempts at composition in this task, to control for the difference in lexical-semantic material preceding the noun in these two conditions, we had participants perform a separate “List” task, also consisting of two-word and one-word trials. In this control task, participants saw either two object-denoting nouns (cup boat) or an unpronounceable consonant string followed by a noun (xkq boat), as before. Importantly, participants were now instructed to decide whether the subsequent colored shape matched any of the words that preceded it rather than all of the words, as before. As the critical stimuli (e.g. boat) were held constant across all four conditions, this control task allowed us to assess whether increases in activity observed during the two-word Composition condition compared to the one-word Composition condition might simply be due to the presentation of two words as opposed to one, as similar differences would be expected to occur within this control task as well. Note that, although noun-noun compounds are grammatical phrases in English, the List task discouraged participants from processing the two-word trials as noun-noun compounds, as they were required to assess the following shape against either one noun or the other. Thus, our design crossed Task (Composition vs. List) and Number of words (one vs. two), with the expectation that only the two-word Composition trials should elicit activity related to linguistic combination.

Each trial contained a fixation, an initial word or non-word, a critical noun, and a target shape. The initial word or non-word varied by condition and could be an
adjective (two-word Composition condition), noun (two-word List condition), or unpronounceable consonant string (one-word conditions). Twenty-five one-syllable nouns were employed in the second (critical) stimulus position (disc, plane, bag, lock, cane, hand, key, shoe, bone, square, bell, boat, bow, car, cross, cup, flag, fork, heart, lamp, leaf, note, star, tree, house). Adjectives in the two-word Composition condition – (red, blue, pink, black, green, brown) – were matched in length with the first stimuli nouns used in the List task – (cup, boat, lamp, plane, cross, house). In both tasks, one-word trials were produced by substituting each adjective or initial noun with an unpronounceable consonant string of the same length – (xkq, qxsw, mtpv, rjdvw, wvcnz, zbxlv). The sets of adjectives, non-words, and first-stimuli List nouns were each created in 28-pt non-proportional Courier font and were all matched in total number of pixels (420.17 avg. [77.39 std.]; 415.12 avg. [49.40 std.]; 420.33 [76.74 std.]; respectively). Target shapes were manually created to depict a canonical, unambiguous representation of one of the nouns, filled in with one of the six colors denoted by the adjectives. Three versions of each target were then created by applying a random scaling factor between 105% and 115% and a random rotation of 0°-360° to the original figure. All stimuli were presented using PsyScope 1.2.5 (J. Cohen et al., 1993) and were projected approximately 50cm from the participant’s eye. Words subtended between 2° and 4° while target shapes were larger, subtending between 6° and 10°.

Both tasks also employed a second, one-word control condition in addition to the unpronounceable consonant strings. This condition used novel symbol strings as the first stimuli (in place of the consonant strings) and was intended to minimize, to
the greatest extent possible, the chance that participants attempted composition between the initial stimulus and the following noun. Thus, for these stimuli, we created repeated symbol strings that carried no obvious linguistic association and were length-matched to the adjectives – (♀♀♀♀, ♂♂♂♂, ☽☾☾, ⃯⃯⃯⃯, ▶▶▶▶, ▽▽▽▽). We then repeated these stimuli as the fixations for all conditions in order to prevent any surprisal or interest related activity when they were substituted for the adjectives. As with the non-words and initial List task nouns, these symbol strings were matched in number of discrete objects, filled area and pixel density to the adjectives used in the two-word Composition condition. However, as the results for this condition were qualitatively identical to the one-word condition described in the text, for simplicity, discussion of this third condition has been eliminated from the main body of the paper.

During each task, participants viewed 300 trials, 100 of each trial type. All conditions contained an equal number of trials in which the target shape matched or did not match the preceding words, while the two-word Composition condition additionally divided the non-matching trials equally among those that did not match the adjective and those that did not match the noun. In these trials, all target shapes matched at least one of the preceding words. During each condition, each of the 25 critical nouns was used four times, twice in matching and twice in non-matching trials. The same set of colored shapes was used as targets for each set of matching and non-matching trials in each condition in each task, such that only 50 different colored shapes were employed in total during each experimental run. However, the trials and stimulus lists were randomized and constructed separately for each participant.
Procedure.

Before the experiment, participants practiced their first task outside of the MEG room. Though participants were made aware of the existence of a second task at this time, no specific instructions regarding the second task were given before the completion of the first task. Instructions and practice for this second task were then given following the completion of the first task, while participants were in the machine. Prior to recording, participants’ head shapes were digitized using a Polhemus Fastrak 3D digitizer (Polhemus, VT, USA). The digitized head shape was then used to constrain source localization during analysis by coregistering five coils located around the face with respect to the MEG sensors.

During the experiment, participants lay in a dimly lit, magnetically shielded room and performed the two tasks in separate blocks. The order of tasks was counterbalanced between participants. MEG data were collected using a using a whole-head 157-channel axial gradiometer system (Kanazawa Institute of Technology, Tokyo, Japan) sampling at 1000Hz with a low-pass filter at 200Hz and a notch filter at 60Hz. During each trial, all stimuli besides the target shapes were presented for 300ms and were followed by a 300ms blank screen. Target shapes appeared at the end of each trial and remained onscreen until the participant made a decision. Subsequent trials began after a blank screen was shown for a variable amount of time. This delay followed a normal distribution with a mean of 400ms and a standard deviation of 100ms. The recording lasted approximately 45 minutes.
Data Acquisition.

MEG data from the 100ms prior to the onset of each critical noun to 500ms post onset were segmented out for each participant for each condition. Raw data were first cleaned of potential artifacts by rejecting trials for which the participant answered either incorrectly or too slowly (defined as over 2.5 seconds after the appearance of the target shape) or for which the maximum amplitude exceeded a threshold which varied between 2500 and 3500fT depending on the amplitude range of each participant. Overall 8.8% (5.6% std.) of trials were excluded. Remaining data were then averaged for each participant for each condition and band-pass filtered between 1 and 40Hz. For inclusion in further analysis, we required that participants show a qualitatively canonical profile of evoked responses during the processing of the critical items. This profile was defined as the appearance of robust and prominent initial visual responses – either the M100 or M170 field pattern (L Pylkkänen and Marantz, 2003; Tarkiainen et al., 1999) – in the time window of 100-200ms following the critical stimuli. In order to assess this criterion, preliminary grand average waveforms were constructed for each participant by averaging over all conditions. Five participants failed to meet this requirement and were excluded from further analysis.

Minimum Norm Estimates.

Distributed minimum norm source estimates served as our primary dependent measure. After preprocessing, a source estimate was constructed for each condition average using L2 minimum norm estimates, calculated in BESA 5.1 (MEGIS Software GmbH, Munich, Germany). The channel noise covariance matrix for each estimate
was based upon the 100ms prior to the onset of the noun in each condition average. Each minimum norm estimate was based on the activity of 1426 regional sources evenly distributed in two shells 10% and 30% below a smoothed standard brain surface. Regional sources in MEG can be regarded as sources with two single dipoles at the same location but with orthogonal orientations. The total activity of each regional source was then computed as the root mean square of the source activities of its two components. Pairs of dipoles at each location were first averaged and then the larger value from each source pair was chosen, creating 713 non-directional sources for which activation could be compared across participants and conditions. Minimum norm images were depth weighted as well as spatio-temporally weighted, using a signal subspace correlation measure (Mosher and Leahy, 1998).

Data Analysis.
To determine effects related to basic combinatorial processing, our primary analysis examined source activity localized to specific regions of interest selected either from studies similar in focus or technique to the present investigation or from historically significant neurophysiological models. To verify the extent and robustness of these results, we also performed a full-brain comparison across all source and time points followed by an analysis of the sensor data targeted at assessing the reproducibility in sensor-space of the significant effects identified by the distributed source analyses.
Regions of Interest.

Regions of interest were drawn upon the smooth cortex used to reconstruct estimated cortical activity from the MEG data (Figure 4). ROIs derived from past fMRI and PET paradigms were given relatively broad interpretations, in order to compensate for the reduced spatial resolution of MEG.

Anterior Temporal Lobe ROIs.

The studies most similar in spirit to the present investigation are those that contrast the processing of sentences to that of non-combinatorial stimuli (e.g. unstructured word lists or environmental sounds). Compared to these baselines, the processing of sentences has consistently shown increased activity in the left anterior temporal lobe (LATL) (Bottini et al., 1994; Friederici et al., 2000; Humphries et al., 2006; Humphries et al., 2005; Humphries et al., 2001; Mazoyer et al., 1993; Stowe et al., 1998; Vandenberghe et al., 2002). The majority of these studies have also identified increased activity within the right anterior temporal lobe (RATL), though nearly always to a lesser degree (Friederici et al., 2000; Humphries et al., 2001; Mazoyer et al., 1993; Stowe et al., 1998). Effects in both hemispheres primarily localize to the anterior temporal pole (approximately BA 38), though many studies have also found increased activity in adjacent regions of the anterior middle and anterior inferior temporal gyri (Bottini et al., 1994; Humphries et al., 2006; Mazoyer et al., 1993; Stowe et al., 1998). Thus, in order to assess the role of the anterior temporal lobes in basic combinatorial processing, we placed an ROI in both the LATL and RATL, each
encompassing roughly the area around BA 38 and the anterior portions of BA 20 and 21 (Figure 4).

Ventro-medial Prefrontal Cortex ROI.
Within MEG, a series of studies has investigated the neural correlates of semantic composition through manipulations involving semantic mismatches (J Brennan and Pylkkänen, 2008; J. Brennan and Pylkkänen, 2010; L Pylkkänen, Martin, et al., 2009; L Pylkkänen and McElree, 2007). In these studies, expressions are contrasted for which the computations necessary in order to construct a coherent meaning are varied while the amount of syntactic work is maximally controlled. The processing of more semantically involved mismatched expressions has consistently elicited increased activity in an MEG component, named the Anterior Midline Field (AMF), that peaks roughly around 400ms and localizes to the ventro-medial prefrontal cortex (vmPFC). Activity in the vmPFC also increases when participants attempt to resolve semantic violations but not when they encounter similar violations of world-knowledge (L Pylkkänen, Oliveri, et al., 2009). The exact location of the vmPFC effect has been somewhat variable across these studies, perhaps due to the difficulty inherent in localizing deep cortical sources in MEG. The probability of correct localization decreases with the depth of the source in MEG (Hillebrand and Barnes, 2002). Thus, there may be both to increased variability and decreased power when measuring activity from a source that is more distant from the sensors, such as the vmPFC. However, the majority of previous effects have localized to the medial region of the orbito-frontal cortex (J. Brennan and Pylkkänen, 2010; L Pylkkänen, Martin, et al.,
2009; L Pylkkänen and McElree, 2007; L Pylkkänen, Oliveri, et al., 2009), with activity occasionally spreading either more ventrally (J. Brennan and Pylkkänen, 2010) or dorsally (J Brennan and Pylkkänen, 2008; L Pylkkänen, Martin, et al., 2009). Thus, in order to assess the contribution of the vmPFC to basic linguistic composition, we placed an ROI within the whole of the medial, ventral portion of the prefrontal cortex (i.e. roughly the entirety of BA 11) (Figure 4).

Traditional Neurolinguistic ROIs.
Historically, by far the most extensively investigated regions in neurolinguistics have been, roughly, the left inferior frontal gyrus (LIFG; “Broca’s area”) and a broad region around the left posterior middle and superior temporal gyri (i.e. the left posterior temporal lobe (LPTL); “Wernicke’s area”). Together these two areas constitute the “traditional” neurophysiological language network. The history of these regions extends well into the nineteenth century (Broca, 1861; Wernicke, 1874) and has yet to coalesce into any resemblance of a consensus regarding their function or even precise location. Despite the paucity of strong evidence directly tying either region to basic linguistic combinatorial processing, we included these familiar regions in our analysis due to their ubiquitous presence throughout neurophysiological studies of language.

The LIFG has frequently been implicated in the processing of many types of complex syntactic constructions such as center-embedding (Caplan et al., 2000; Stromswold et al., 1996), wh-extraction (Santi and Grodzinsky, 2007), and scrambled verbal arguments (Friederici et al., 2006). Along with these syntactic findings, the
LIFG has also been implicated in many processes that are not strictly syntactic in nature, for example the maintenance and manipulation of working memory (J. D. Cohen et al., 1997; Fiebach et al., 2007), cognitive control (Badre and Wagner, 2007), and executive functioning (Koechlin and Summerfield, 2007). Due to the breadth of this empirical coverage, the LIFG has been suggested to subserve any number of mental mechanisms from the narrowly specific (e.g. syntactic movement only; Grodzinsky and Santi, 2008) to the broadly general (e.g. selection and regulation among competing mental representations; Thompson-Schill et al., 2005). While the exact placement of the LIFG has varied alongside its multi-faceted interpretations, in general the pars opercularis and pars triangularis (approximately BA 44 and 45) have served as the center for these investigations. Therefore, we created a relatively broad ROI around these regions from which to measure LIFG activity in the present study (Figure 4).

The LPTL region has become decidedly more fragmented over the years, dividing into several apparently functionally distinct areas. The posterior portion of the middle temporal gyrus (pMTG) has been most heavily implicated in lexical-level operations (i.e. the storage and retrieval of long-term linguistic information). Activity measured by fMRI within the pMTG exhibits lexical priming effects across a wide range of stimulus-onset asynchronies and modalities (Gold et al., 2006) and has been found to increase with the number of words processed per trial (Badre et al., 2005) and for judgments requiring lexical-semantic information (Gitelman et al., 2005). Within MEG, activity localized to this general region has shown sensitivity to many factors that affect the timing of lexical access, such as frequency (Embick et al., 2001),
repetition (L. Pylkkänen et al., 2000), and phonotactic probability (L. Pylkkänen et al., 2002) (see L. Pylkkänen and Marantz, 2003; Salmelin, 2007 for reviews). Thus, there is converging evidence from both the hemodynamic and electrophysiological literature that the pMTG plays an integral role in the storage and retrieval of lexical information.

In relation to combinatorial operations, it has been suggested that classic N400 effects, canonically elicited by semantically incongruous sentence completions, e.g. *He spread the warm bread with socks* (Hagoort et al., 2004; Kutas and Hillyard, 1980), localize, at least partially, to the pMTG (see Lau et al., 2008 for a review). However, the exact relationship between these N400 effects and combinatorial processing is currently unclear, as the N400 is also modulated by many lexical-level factors, such as frequency (Allen et al., 2003; van Petten and Kutas, 1990) and repetition (Rugg, 1985), and can be elicited within non-compositional contexts, such as semantically unrelated word pairs (e.g. Holcomb, 1993).

Similar ambiguity surrounds an adjacent region of the LPTL, the angular gyrus (AG), located superior and posterior to the pMTG (roughly BA39). The AG has recently been suggested to play a role in the integration of semantic information into context (Lau et al., 2008), as increased AG activity has been observed in response to semantically congruous sentences compared to semantically related word lists or syntactically well-formed but meaningless sentences (Humphries et al., 2006, 2007). Also, the AG has occasionally exhibited more activity during the reading of sentences compared to word lists (Bavelier et al., 1997; Bottini et al., 1994). However, the AG has been implicated in lexical-level processing as well, showing increased activity during the presentation of single words compared to non-words during both lexical
decision tasks (Ischebeck et al., 2004; Rissman et al., 2003) and simple reading (Binder et al., 2005; Bookheimer et al., 1995).

Despite the fractionation and uncertainty surrounding the LPTL, for the sake of simplicity, we enclosed both the pMTG and the AG within a single LPTL ROI in order to identify any composition related activity in this region (Figure 4). (However, it should be briefly noted here that dividing this region into two distinct ROIs does not substantively affect the results.)

ROI Analysis.

Significant effects within each ROI were determined by applying a non-parametric, cluster-based permutation test (Maris and Oostenveld, 2007) to the entire time interval surrounding the onset of the critical noun (0-500ms). For this analysis, we constructed a test statistic capable of isolating clusters of time points for which the activity localized to each ROI exhibited the pattern we hypothesized for composition related activity. Specifically, we identified periods of contiguous time points for which the increase in activity during the two-word Composition condition compared to the one-word Composition condition was significantly greater than any activity difference seen between the two List task conditions. ROIs exhibiting such effects were then further investigated using a follow-up permutation test within each task. This follow-up test employed a similar but simpler test statistic applicable for identifying increases in one condition relative to another.

The logic for the permutation test is as follows. First, a cluster-based test statistic for the observed data is calculated (as described below). Then, the same test statistic is
calculated for many permutations of the actual data, creating a distribution for the statistic. Permutated data sets are created by randomly assigning the condition labels within each participant. Importantly, this assignment is done independently for each participant and only once per participant per permutation (as opposed to once for each time point for each participant). The $p$-value of the observed test statistic is then computed relative to this distribution and is set equal to the proportion of permuted data sets that produce a test statistic more extreme than that of the actual data.

Our experimental design was constructed such that composition related activity should exhibit a particular type of interaction between Task and Number of words. Specifically, within the Composition task, we expected composition related activity to be greater within the two-word condition compared to the one-word condition, while within the List task we expected to see no composition related activity in either condition. Thus, following the discussion in (Maris and Oostenveld, 2007, p. 187), we employed a test statistic for our permutation test that incorporates this prior hypothesis and is therefore capable of identifying these particular types of interactions. First, we identified clusters of adjacent time points for which there was an interaction between Task and Number of words by submitting the data at each time point to a $2 \times 2$ repeated-measures ANOVA with Task (Composition vs. List) and Number of words (one vs. two) as factors. Clusters were then extracted by identifying intervals of at least 10 adjacent time points for which, at each time point, the resulting $p$-value for the interaction between the two factors was less than a set threshold, chosen so as to identify relatively shallow yet long-lasting patterns of activity ($p = 0.30$). Then, at each time point in a given cluster, we performed a paired t-test within each task. This
gave us two $t$-values for every time point in the cluster – one corresponding to the difference in activity between the two Composition conditions and one corresponding to the difference in activity between the List conditions. The test statistic for a cluster was then calculated by summing all Composition $t$-values and subtracting the magnitude of all List $t$-values. The use of this test statistic allowed us to specifically identify clusters for which the interaction observed in the initial phase of the test was driven by two concurrent factors: (1) an increase in activity during the two-word Composition condition relative to the one-word Composition condition and (2) no difference in activity between the List task conditions. In other words, this test statistic, and consequently the test itself, incorporates our prior hypothesis regarding the particular type of interaction that we are interested in, and so can uncover effects directly relevant to this hypothesis.

For follow-up permutation tests within each task, clusters were identified by performing a paired $t$-test between the one-word and two-word activity measures at each time point and then extracting intervals for which the resulting $p$-value was less than the same set threshold as before ($p = 0.30$) for a minimum of 10 adjacent time points. The test statistic for each cluster was then calculated simply as the sum of each $t$-value for each time point in the cluster. For all tests, the reported $p$-value was obtained from a distribution of test statics derived from 10,000 permutations of the original data.

Whole brain analysis.

Whole brain analyses were performed within each task. Two-word and one-word
activity measures were compared sample by sample for every source-time point using a paired \( t \)-test. A difference was considered significant if it remained reliable \( (p < 0.05) \) for at least 10 samples (10ms) and was observed in at least 10 adjacent cortical sources. In the results and figures below, we discuss only effects due to an increase in two-word activity compared to one-word activity within each task.

*Sensor space analysis.*

Our sensor space analysis was designed to determine the extent to which effects identified through the source space analyses were also apparent in the sensor data. Thus we performed one sensor space analysis for each significant, combinatorial source space effect as determined by the ROI analyses. For each such effect, we first identified a set of sensors of interest for each participant based on the spatio-temporal location of the source space result. Then, we performed the same cluster analysis as for the source space ROIs on the RMS of the data from these sensors.

Since the mapping between source space and sensor space was not clear a priori, we adopted a split-half, cross-validation method to analyze the sensor data. Initially, the data was divided into two halves by randomly partitioning the trials used for source reconstruction, for each participant for each condition. We then reaveraged each data partition, producing two equally sized sets of data, each containing an average per participant per condition. As during source reconstruction, these averages were filtered between 1Hz and 40Hz for analysis. Sensors of interest for each participant were then selected based upon the data from the two-word Composition condition. In one half of the data, the training set, we first visually inspected the grand
average sensor data for the two-word Composition condition, accumulated from all of the participants, and identified the dominant field pattern present in this data during the peak of the source space effect. Then, for each participant, we inspected their individual two-word Composition condition average, also from the training set, and identified the set of sensors whose data most closely resembled this dominant component. In each case, we chose the sensors in the centers of the ingoing and outgoing fields of the dominant component and the set of sensors surrounding these peaks for which the field pattern remained strongly present. Across all selections, this resulted in an average of 12.95 sensors per set (min: 6; max: 25; std: 4.065). If the dominant component was not apparent in the participant’s average, that participant was excluded from the current analysis. To create the dependent measure for input into the cluster tests, we then calculated the RMS for the set of sensors chosen for each participant using the other half of the data, the testing set. Thus, the data used to select the sensors was independent of the data used in the analysis. We applied the cluster test to this data over the entire interval (0-500ms) as before and also to a more circumscribed time interval surrounding the extent of the corresponding source space effect, in order to compensate for the loss of power that results from partitioning the data. Finally, we switched the training and test sets and repeated the analysis for each source space effect. Thus, ultimately, we produced two complementary sensor space analyses for each source space effect of interest.
Results

*Behavioral Results.*

Accuracy and reaction time data for responses were submitted to a 2 x 2 repeated-measures ANOVA with Task (Composition vs. List) and Number of words (one vs. two) as factors (Figure 2). There were no significant effects found for accuracy, either for Task, Number of words, or their interaction (all $F$s < 1), as participants were generally close to ceiling on all conditions (two-word Composition: 96.9% avg. [3.2% std.]; one-word Composition: 96.9% avg. [2.9% std.]; two-word List: 96.6% avg. [4.2% std.]; one-word List: 97.5% avg. [3.0% std.]).

We found a strong interaction between Task and Number of words for reaction times, ($F$(1,19) = 89.2, $p < 0.001$). Paired $t$-tests within each task revealed that, compared to the corresponding one-word conditions, participants were markedly slower for the two-word condition within the List task (762ms avg. [173ms std.] vs. 679ms avg. [173ms std.]; $p < 0.001$) while they were significantly faster for the two-word condition in the Composition task (609ms avg. [165ms std.] vs. 658ms avg. [186ms std.]; $p < 0.001$). These results indicate that, in the List task, the two-word condition was more difficult than the one-word condition, as participants were much slower to respond and were slightly less accurate in the former compared to the latter. As participants needed to remember twice as many items in the two-word condition compared to the one-word condition, this result is relatively unsurprising and echoes many past studies that demonstrate response times in memory tasks increase with the number of items to remember (e.g. Sternberg, 1967). Conversely, in the Composition
task, the two-word condition appeared to be the easier of the two, as participants responded significantly faster in this condition than in the one-word condition. This result is somewhat surprising, as participants had to assess the correspondence of twice as many features in the two-word condition compared to the one-word condition at the presentation of the target shape. However, this effect is in line with previous results demonstrating that, following an adjective-noun description, participants could identify pictures that incorporated the adjective faster than those that did not (Potter and Faulconer, 1979). Importantly, the present behavioral results suggest that any increase in neural activity observed for the two-word Composition condition does not reflect increased effort, as this critical condition appeared to be the easiest of the four.

*General Assessment of MEG Sensor Data.*

A qualitative overview of the evoked MEG response during the presentation of the critical nouns is shown in Figure 3. The initial baseline activity preceding the onset of the noun (used to estimate the noise covariance matrix during the construction of each minimum norm estimate) showed no significant differences in average amplitude between conditions (Task x Words two-way repeated-measures ANOVA yielded all $F$s < 0.5; two-word Composition: 16.09 fT avg. [3.89 fT std.]; one-word Composition: 16.60 fT avg. [4.86 fT std.]; two-word List: 16.65 fT avg. [5.72 fT std.]; one-word List: 16.62 fT avg. [4.52 fT std.]). Across all conditions, we observed canonical early visual responses at approximately 100ms and 150ms that have consistently been identified in MEG during the presentation of visual words (L Pylkkänen and Marantz, 2003; Tarkiainen et al., 1999). These visual responses were followed by the
characteristic M250 and M350 field patterns, focused over the left temporal lobe, that also have been consistently observed following visually presented words (Embick et al., 2001; L Pylkkänen and Marantz, 2003; L. Pylkkänen et al., 2002). Additionally, a large, sustained increase in MEG activity can be seen in the two-word Composition condition beginning at approximately 350ms and continuing to 450ms. This activity is accompanied by the characteristic AMF field pattern (e.g. L Pylkkänen and McElree, 2007), which we did not observe in any of the control conditions during this time.

**Anterior Temporal Lobe ROIs.**

Significant clusters of time points were identified by the interaction permutation test within both the LATL and RATL (Figure 4). Within the LATL, a significant cluster of combinatorial activity was found from **184ms-255ms** ($p = 0.0389$; two-word Composition: 3.87 nAm avg. [1.74 nAm std.]; one-word Composition: 2.85 nAm avg. [0.97 nAm std.]; two-word List: 3.29 nAm avg. [1.04 nAm std.]; one-word List: 3.31 nAm avg. [1.11 nAm std.]). A follow-up test within the Composition task alone revealed a temporally similar cluster of time points for which activity increased during the two-word condition compared to the one-word condition, though this cluster was only marginally significant (**204-263ms**; $p = 0.071$; two-word: 4.13 nAm avg. [1.97 nAm std.]; one-word: 3.03 nAm avg. [1.20 nAm std.]). Within the List task, no clusters of increased activity during the two-word condition were seen at any time within the LATL (all clusters $p > 0.35$). Inspecting the waveforms from this ROI, there appears to be an additional increase in two-word Composition activity relative to the other three conditions peaking at 375ms (Figure 4). However, despite the visual
clarity of this separation, activity during this time did not approach significance in
the interaction cluster test (closest cluster: 327-334; \(p = 0.5549\)). A follow-up test within
the Composition task alone also failed to approach significance (closest cluster: 360-
380; \(p = 0.4218\)). Even a direct comparison of two-word to one-word Composition
activity levels from 350-400ms using a one-tailed paired \(t\)-test only approached
significance (\(p = 0.076\); two-word: 3.75 nAm avg. [1.73 nAm std.]; one-word: 3.13
nAm avg. [1.56 nAm std.]). Thus, despite the tantalizing nature of this visual
separation, no firm conclusions can be drawn at this time about activity within the
LATL during this later time window. There is, however, strong evidence that activity
in the LATL reflects by basic combinatorial processing in the earlier time window,
from 184 to 255ms.

Within the RATL, two highly significant clusters were identified by the initial
permutation test – 184-246ms (\(p = 0.0078\); two-word Composition: 3.86 nAm avg.
[1.72 nAm std.]; one-word Composition: 2.83 nAm avg. [1.13 nAm std.]; two-word
List: 3.73 nAm avg. [1.34 nAm std.]; one-word List: 3.57 nAm avg. [1.58 nAm std.])
and 329-403ms (\(p = 0.0041\); two-word Composition: 4.26 nAm avg. [2.02 nAm std.];
one-word Composition: 2.86 nAm avg. [1.12 nAm std.]; two-word List: 3.69 nAm
avg. [1.53 nAm std.]; one-word List: 3.67 nAm avg. [1.93 nAm std.]). A follow-up
test within the Composition task also identified two clusters of significantly greater
activity in the two-word condition compared to the one-word condition at
approximately the same times – 163-270ms (\(p = 0.0168\); two-word: 4.03 nAm avg.
[1.65 nAm std.]; one-word: 2.92 nAm avg. [1.12 nAm std.]) and 328-410ms (\(p =
0.0248\); two-word: 3.89 nAm avg. [1.94 nAm std.]; one-word: 2.70 nAm avg. [1.17
nAm std.]). Within the List task, no significant clusters of increased two-word activity were identified at any time in the RATL (all clusters $p > 0.40$).

Though these results certainly suggest a role for the RATL within the present manipulation, the exact relation between these effects and combinatorial processing is not quite as clear as in the LATL. Within the LATL, activity for the two-word Composition condition increased relative to the other three conditions during the identified cluster. Within the RATL, on the other hand, the one-word Composition condition appears to be the outlier, falling well below the other three conditions for the majority of the interval. While it is difficult to draw firm conclusions comparing across tasks (as they were performed sequentially and not concurrently), it seems more likely that composition related activity would display the pattern seen within the LATL (i.e. increased activity in the two-word Composition condition relative to the other three conditions). Consequently, the profile within the RATL seems to suggest a suppression of activity within the one-word Composition condition rather than an increase in the two-word condition. This interpretation is strengthened by the fact that, in our paradigm, across both tasks the two one-word conditions were visually identical and ostensibly required the same computational demands (i.e. comparing the colored shape to the single object-denoting noun that preceded it). Therefore, the difference in activity levels observed within the RATL suggests a task related decrease in activity during the one-word Composition condition, though it is unclear exactly which aspect of the task might engender this response.

Generally, our results suggest that activity in both the LATL and RATL is modulated by basic combinatory processes. Specifically, the overall profiles of the two
activities suggest that the RATL is subject to a task-related suppression of activity during the one-word Composition condition while the LATL reflects basic combinatory operations around 225ms following the presentation of a composable noun.

*Ventro-medial Prefrontal Cortex ROI.*

Within the vmPFC ROI, the interaction permutation test identified a highly significant, long-lasting cluster of time points for which localized activity exhibited the combinatorial profile (331-480ms; \( p = 0.0135 \); two-word Composition: 3.61 nAm avg. [2.59 nAm std.]; one-word Composition: 2.74 nAm avg. [1.57 nAm std.]; two-word List: 2.69 nAm avg. [0.95 nAm std.]; one-word List: 3.02 nAm avg. [1.71 nAm std.]; see Figure 4). A follow-up test within the Composition task again conformed to this result, identifying increased activity in the two-word condition compared to the one-word condition from 326-442ms (\( p = 0.0263 \); two-word: 3.76 nAm avg. [2.95 nAm std.]; one-word: 2.78 nAm avg. [1.64 nAm std.]). Within the List task, no clusters of increased activity for the two-word condition were identified at any point (all clusters \( p > 0.85 \)). Overall, the profile of activity within the vmPFC resembled that observed within the LATL. Activity in the two-word Composition condition was clearly greater than in the other three control conditions. This result suggests that activity localized to the vmPFC reflects basic linguistic combinatorial operations approximately 300-500ms following the onset of a composable noun.
Lateral Inferior Frontal Gyrus and Posterior Temporal Lobe ROIs. Within both the LIFG and the LPTL ROIs, the interaction permutation test failed to reveal any significant clusters of combinatorial activity (all clusters $p = 1$; see Figure 4). A targeted test within the Composition task alone also failed to produce any clusters for which activity was significantly greater in the two-word condition compared to the one word condition in either region (LIFG: all clusters $p > 0.60$; LPTL: all clusters $p > 0.30$). Within the List task, the LIFG showed no significant increases in two-word activity (all clusters $p > 0.40$) while the LPTL did exhibit one cluster in which there was a slight trend towards increased activity in the two-word condition (206-266ms; $p = 0.1430$; two-word: 2.82 nAm avg. [1.66 nAm std.]; one-word: 2.17 nAm avg. [1.13 nAm std.]). This cluster is discussed within the context of the full-brain analysis below.

A division of the LPTL into two ROIs – one encompassing roughly the pMTG and one centered within the AG – also failed to produce any significant clusters of combinatorial activity (pMTG: all clusters $p = 1$; AG: all clusters $p > 0.20$). However, results from post-hoc tests within each task do begin to reveal an interesting hint of a dichotomy between these two areas. Activity within the AG demonstrated a trend towards increased two-word activity within the Composition task (144-213ms; $p = 0.1183$; two-word: 2.60 nAm avg. [1.40 nAm std.]; one-word: 2.01 nAm avg. [1.01 nAm std.]) but not within the List task (all clusters $p > 0.30$). The pMTG, on the other hand, exhibited no combinatorial increases within the Composition task (all clusters $p > 0.55$) while within the List task several clusters trended towards increased activity in the two-word condition (107-175ms; $p = 0.1141$; two-word: 2.72 nAm avg. [1.30 nAm std.]).
Thus, our results potentially implicate the AG during combinatorial processing, while the pMTG seems to play more of a role within the List task. However, due to the extremely tentative nature of these effects, we refrain from further speculation.

In general, the results from these two areas suggest that past effects observed within the LIFG and LPTL primarily reflect either non-combinatorial or more complex combinatorial mechanisms within language processing. However, it is, of course, also possible either that our paradigm (i.e. MEG in conjunction with our analysis method) is ill-suited to reproduce past results in these regions or that the present manipulation is too subtle to adequately bring out combinatorial effects in these regions.

**Full Brain Comparisons.**

In general, our full brain analyses (Figure 5) conform quite closely to our ROI analyses. Within the Composition task, a clear LATL effect can be seen around 225ms, followed by an increase in two-word activity centered in the vmPFC at approximately 400ms. Throughout, a sustained difference can also be seen in the RATL, mirroring our earlier ROI analysis. No obvious effects appear in either the LIFG or LPTL ROIs.

In the Composition task, the only other significant effect of note is a sustained increase in activity during the two-word condition localized to a large region of the left parietal lobe, roughly near the central sulcus and superior parietal lobule (SPL).
The central sulcus, of course, houses the primary motor cortices, and activity in this region has been shown to occur not only in overt speech but covert speech as well (Huang et al., 2002; Pulvermüller et al., 2006; Riecker et al., 2000). Thus, it is possible that participants were disproportionately engaged in covert speech during compositional processing compared to non-compositional processing. Another, potentially more intriguing hypothesis arises from past work implicating the SPL in the binding of visual features. Lesions in the SPL have been associated with deficits in correctly combining the visual features of an object (Friedman-Hill et al., 1995) and also deficits in visual search tasks requiring the conjunction of features (Robertson et al., 1997). Further, increased activity in this region has been observed during conjunction searches compared to those that target only one visual feature (Corbetta et al., 1995). Thus it is possible that the increase in activity we observe near the SPL is related in some manner to the feature binding required by the two-word Composition condition, though the exact nature of this link is unclear at the present time. It should be noted, however, that there is much debate surrounding the role of the SPL in visual feature binding. Many studies suggest that effects in this region during visual conjunction search are primarily due not to the binding of features but to the deployment of spatial attention (Donner et al., 2002; Nobre et al., 2003; Shafritz et al., 2002), argued to be a necessary precursor to visual feature binding (Treisman and Gelade, 1980). Within the context of this hypothesis, the present result might be related to a preemptive deployment of attention in preparation for the upcoming target shape in the two-word Composition condition, as participants are aware binding will be needed for their forthcoming decision given the preceding adjective. However,
while parallels between the present task and past visual search studies are intriguing, hypotheses relating the present results to visual attention or feature binding can only be speculative at the present time.

Within the List task, increased activity for the two-word compared to the one-word condition can be seen within the inferior, posterior portion of the left temporal lobe and within the right inferior frontal gyrus (RIFG) at approximately 200-300ms. The RIFG has long been implicated in verbal working memory (J. D. Cohen et al., 1994; Fiez et al., 1996; Petrides et al., 1993; Wei et al., 2004). Thus, it is not surprising to find increased activity within this region during the present task when participants had to remember two words compared to one. However, not only have past effects generally been found bilaterally, but activity within the LIFG is usually greater than in the RIFG during memory tasks (Awh et al., 1996; Paulesu et al., 1993; Rypma et al., 1999). Thus, the appearance of only a right hemisphere effect within the present contrast is somewhat unexpected. The left temporal effect, which drives the trend towards increased two-word List activity observed in the LPTL ROI, is also somewhat unexpected, as this area has not canonically been linked to verbal memory tasks. However, increased activity in this general region has been reported during non-verbal memory tasks (Courtney et al., 1996; Ungerleider et al., 1998).

Compared to past experiments, the manipulation we employed in the List task is relatively weak with respect to memory demands. Thus, the effects that we observed might only make up a muted and incomplete snapshot of the full network of neural regions that participate in this task. Importantly, however, the manipulation within the List task with respect to the number of lexical items in each condition is equivalent to
that in the Composition task. Therefore, any effects seen within this critical task due
merely to the difference in lexical material between conditions should have been
observable within the List contrast as well. The only hint of such shared activity is
within the RIFG, as increased activity in this region can be seen from 200-300ms in
the Composition contrast as well. However, no vmPFC or LATL effects are visible at
all within the full-brain List comparison. This suggests that the effects observed within
these regions during the Composition task are not due simply to a difference in the
number of lexical items.

Sensor Space Analysis.

The two clearest combinatorial effects resulting from our source space analysis were
in the LATL from approximately 175-275ms and in the vmPFC from approximately
300-500ms. Thus, we performed two separate sensor space analyses targeted at further
characterizing the relationship between these source space effects and the recorded
sensor data.

Left Anterior Sensors:

Both split-half analyses revealed a robust combinatorial effect within the left
hemisphere sensors during the same time window as the LATL source space effect,
approximately 175-275ms, (see Figure 6). Across both halves, the dominant field
pattern observed within the grand average of the two-word Composition condition at
this time and location could be identified in 18 of the 20 participants, however, in each
half a different set of two participants failed to show this component. In the first half,
within the targeted time window from 175-275ms, the interaction permutation test revealed a significant cluster of combinatorial activity from \textbf{195-215ms} ($p = 0.036$; two-word Composition: 48.24 fT avg. [40.60 fT std.]; one-word Composition: 28.26 fT avg. [18.58 fT std.]; two-word List: 35.15 fT avg. [31.78 fT std.]; one-word List: 31.80 fT avg. [28.27 fT std.]). A follow-up test within the Composition task identified a significant cluster of increased activity in the two-word condition compared to the one-word condition from \textbf{196-235ms} ($p = 0.037$; two-word: 46.78 fT avg. [39.35 fT std.]; one-word: 30.65 fT avg. [20.66 fT std.]). No such clusters were found within the List task (all clusters $p > 0.25$). In the expanded time window of 0-500ms, the significance level of these clusters was reduced, reflecting the loss in power due to partitioning the data (Interaction: $p = 0.19$; Composition: $p = 0.17$). In the second data half, the targeted permutation tests again found both a significant cluster of combinatorial activity using the interaction test (\textbf{220-241ms}; $p = 0.013$; two-word Composition: 34.89 fT avg. [20.10 fT std.]; one-word Composition: 21.17 fT avg. [12.91 fT std.]; two-word List: 27.12 fT avg. [19.22 fT std.]; one-word List: 24.09 fT avg. [21.72 fT std.]), and a significant cluster of increased two-word activity in the Composition task (\textbf{218-255ms} ($p = 0.018$; two-word: 32.41 fT avg. [17.62 fT std.]; one-word: 20.28 fT avg. [11.19 fT std.])). No significant effects were identified in the List task (all clusters $p > 0.8$). In the expanded time window, the significance of the identified clusters was again somewhat reduced (Interaction: $p = 0.08$; Composition: $p = 0.12$).
Anterior Midline Sensors:

Both split-half analyses also revealed a robust combinatorial effect within the frontal sensors during the later time window surrounding the identified vmPFC source effect, 300-500ms, (see Figure 6). In the first test set, three participants failed to exhibit the dominant frontal field pattern, while in the second set, two participants were excluded, with one participant overlapping between the two exclusion sets. In the first split-half analysis, a targeted interaction permutation test, from 300-500ms revealed a significant cluster of combinatorial activity from 396-448ms ($p = 0.011$; two-word Composition: 27.10 fT avg. [25.12 fT std.]; one-word Composition: 17.12 fT avg. [15.64 fT std.]; two-word List: 17.68 fT avg. [9.18 fT std.]; one-word List: 18.35 fT avg. [10.84 fT std.]). A follow-up permutation test within the Composition task showed a significant cluster of increased two-word activity from 371-448ms ($p = 0.012$; two-word: 25.51 fT avg. [21.77 fT std.]; one-word: 16.00 fT avg. [13.77 fT std.]). No such clusters were identified within the List task (all clusters $p > 0.40$). The relaxed time window from 0-500ms did not qualitatively reduce the significance of these results (Interaction: $p = 0.027$; Composition: $p = 0.030$). In the second data half, the targeted permutation tests revealed a significant cluster of combinatorial activity from 388-442ms ($p = 0.050$; two-word Composition: 29.20 fT avg. [20.01 fT std.]; one-word Composition: 21.24 fT avg. [13.80 fT std.]; two-word List: 20.11 fT avg. [12.65 fT std.]; one-word List: 24.80 fT avg. [17.43 fT std.]) and a significant cluster of increased two-word activity within the Composition task from 389-439ms ($p = 0.033$; two-word: 29.38 fT avg. [20.51 fT std.]; one-word: 21.02 fT avg. [13.65 fT std.]). No significant cluster of increased two-word activity was found within the List
task (all clusters $p > 0.85$). This time, expanding the time window slightly reduced the significance values of the two identified clusters (Interaction: $p = 0.11$; Composition: $p = 0.15$).

Thus, the results of our sensor space analysis indicate that both the LATL and vmPFC source space effects are reflected rather robustly in sensor space as well. A significant combinatorial effect was found in the left hemisphere sensors in the same time range as the LATL effect, approximately 175-275ms. Likewise, an analysis of the frontal sensors revealed a significant cluster of combinatorial activity in the same time window as the vmPFC effect, approximately 300-500ms. The dominant field pattern associated with this latter effect appears to be the AMF (see Figure 6), thus replicating previous findings relating this field pattern to vmPFC source activity during the resolution of semantic mismatches (e.g. L Pylkkänen and McElree, 2007).
Discussion

The present study deviates from the neurolinguistic norm of treating basic combinatorial processing as the baseline against which more complex operations are measured. Instead, we explicitly investigated the most minimal of combinatorial processes – the simple composition of an adjective with a noun. Using MEG, we isolated neural activity elicited by the comprehension of a simple, object-denoting noun (*boat*) and measured increases in this activity evoked within a minimal compositional context (*red boat*) compared to a matched control (*xkq boat*). By contrasting this manipulation to that of a similar non-compositional context (*cup, boat*) compared to the same control, we identified activity associated with basic composition that cannot be attributed simply to an increase in lexical-semantic material. Though several past studies have compared combinatorial and non-combinatorial linguistic processing, none have reduced the scope of their investigation to such a minimal and fundamental level. Further, as these past studies have used either fMRI or PET, the present results provide a unique temporal map of these critical operations.

*ATL and vmPFC Effects of Composition.*

Both the LATL and vmPFC demonstrated significant combinatorial activity during our minimal manipulation. In both regions, we found clear clusters of time points for which activity during the processing of the nouns was significantly greater during the compositional context compared to the matched control. The LATL effect peaked at
approximately 225ms and was followed by increased vmPFC activity at 400ms. Importantly, no corresponding increases were observed in the non-compositional contrast. The RATL also exhibited a clear difference in activity during the compositional manipulation, however, the activity profile in this region instead suggests a task-related suppression during the control condition, as activity in this condition was significantly lower than in the other, structurally identical one-word control.

Both the LATL and RATL have previously been implicated in sentence processing. The comprehension of sentences consistently produces increased activity in these regions compared to unstructured word lists (e.g. Friederici et al., 2000; Mazoyer et al., 1993; Stowe et al., 1998). Thus, the present results suggest that very basic combinatorial operations drive the LATL effects observed in these manipulations. Regarding the RATL, our results are more ambivalent and introduce the possibility that past “increases” observed within this region might in fact reflect decreases in neural activity during non-compositional processing. Though variability in experimental design makes direct comparisons difficult, the markedly different activity profiles exhibited by the LATL and RATL in the present results may help to illuminate past asymmetries found between these two regions. In previous studies, RATL effects are consistently smaller than LATL effects (e.g. Mazoyer et al., 1993; Stowe et al., 1998) and occasionally not present at all (Bottini et al., 1994; Vandenberghe et al., 2002). Whether the dichotomy suggested here – that the LATL subserves basic linguistic composition while the RATL undergoes a task-related suppression – can explain these differences will require more work.
The vmPFC has recently been implicated as important in resolving semantic mismatches (e.g. J Brennan and Pylkkänen, 2008; L Pylkkänen and McElree, 2007). Several studies have compared compositionally transparent constructions to syntactically matched controls that require more work in order to obtain a coherent meaning. Results have consistently implicated the vmPFC as important in resolving the more semantically intensive expressions. The present findings, therefore, suggest that the vmPFC supports the construction of basic linguistic meaning and that past effects were not driven by more specialized mechanisms related to semantic mismatch resolution in particular.

This hypothesis – that the vmPFC reflects basic semantic composition – raises an intriguing possibility within the context of the present results. All linguistic combination, even that associated with basic adjective-noun phrases, can be broken down into two broad types of combinatorial processing, the construction of phrases based on grammatical categories – syntactic composition – and the creation of complex meanings from simpler pieces – semantic composition. In contrast to the vmPFC, the LATL has been most strongly associated with syntactic combinatoric processing. Activity in the LATL correlates with measures of syntactic complexity during natural story comprehension (J. Brennan, Nir Y. et al., 2010) and has been found to exhibit reduced activity in contexts that elicit syntactic priming (Noppeney and Price, 2004). Therefore, though our design did not aim to explicitly disentangle these two types of combinatorial processes, the present results are consistent with the interpretation that the LATL subserves basic syntactic combinatorial operations while the vmPFC supports basic semantic composition. Furthermore, a broad range of
processing models mirror the temporal ordering of our results and place the initiation of syntactic composition prior to semantic composition (e.g. Friederici, 2002).

Of course, neither this hypothesis nor our results imply that the LATL and vmPFC solely perform combinatorial operations. Both regions have also been implicated in non-combinatorial, lexical-level tasks as well, such as lexical decision (Fujimaki et al., 2009; Mummery et al., 1999; Nobre et al., 1994), repetition detection (Halgren et al., 1994), and semantic relatedness judgments (Vandenberghe et al., 1996). Additionally, both regions exhibit classical lexical–level effects such as decreased activity for more frequent (Halgren et al., 2002) or primed words (Dale et al., 2000; Dhond et al., 2001; Marinkovic et al., 2003; Mummery et al., 1999). However, in the present study, lexical computations should be relatively constant across both tasks. The same words served as critical items in each condition, and the probability of a particular critical item following any preceding item was equated across all stimulus pairs within the context of the experiment. Therefore, while our results clearly do not deny a role for the vmPFC or LATL in non-combinatorial, lexical level processing, it is not immediately apparent how such operations would produce the effects observed in the present experiment.

It should be emphasized that the partitioning of syntactic and semantic combinatorial processes to the LATL and vmPFC can at best be tentative at the present time. In general, disentangling syntactic and semantic processes is a very difficult problem, with disagreement on even the extent to which a solution is theoretically possible (see L Pylkkänen, 2008). The majority of studies investigating compositional processes, including the present one, manipulate syntax and semantics
in tandem, thus making it difficult to assign either one process or the other to observed effects, and past attempts to disentangle the two have not always produced converging evidence. In particular, while much evidence implicates the LATL in syntactic combinatorial processing, a number of researchers have also suggested that this region plays an important role in semantic combinatorial operations as well. For example, when participants are asked to monitor sentences for semantic anomalies a subset of the LATL shows increased activity compared to monitoring for syntactic violations (Rogalsky and Hickok, 2009). Also, scrambling the content words of sentences modulates activity in the LATL, though the direction of this effect has varied, with scrambled sentences eliciting both more (Vandenberghe et al., 2002) and less (Humphries et al., 2006) LATL activity compared to normal sentences. Clearly, much work is still needed before any conclusions can be solidly established regarding the apportionment of syntactic and semantic combinatorial operations to specific cortical regions, however, the present study does provide a novel, minimal framework upon which to build, and the proposed delineation provides an intriguing point of departure for future studies.

_Lack of LIFG and LPTL Effects._

Neither the LIFG nor LPTL exhibited any significant activity related to basic combinatorial processing. Of course, as with any null result, the absence of these effects may be due to any number of factors such as the particular technique used (the majority of past studies implicating these areas in language processing were conducted within fMRI while the present study localizes MEG data), the analysis method
(perhaps a more nuanced localization method might provide more power), or the subtlety of the manipulation (the simple act of composing an adjective with a noun might not be enough to drive certain types of combinatorial activity sufficiently above the background noise). However, these null results are also consistent with past evidence suggesting that these regions support either relatively complex or non-combinatorial (i.e. lexical-level) linguistic operations.

Within the LPTL, the pMTG has primarily been linked to long-term lexical storage and access (Gold et al., 2006; L. Pylkkänen et al., 2002). If this interpretation is correct, then one would not expect to see effects in this region during the present manipulation, since, as discussed above, lexical computations should be constant across tasks. The more posterior portion of the LPTL, the AG, has shown increased activity for semantically coherent sentences compared to syntactically well-formed but incoherent sentences (Humphries et al., 2007). However this region, like the pMTG, has also primarily been associated with lexical-level tasks (e.g. Binder et al., 2005; Rissman et al., 2003). Thus, the lack of any strong combinatorial effects within the LPTL in the present study is consistent with this region subserving either lexical-level operations or more complex aspects of sentence processing not elicited by the present manipulation.

The LIFG has canonically been associated with the processing of complex syntactic structures (e.g. Stromswold et al., 1996) and high-level cognitive functioning (e.g. Badre and Wagner, 2007). Thus, we do not find the lack of combinatorial effects within this region particularly surprising given the present minimal manipulation. This outcome would be predicted by any number of hypotheses posited for this region, e.g.
syntactic movement (Grodzinsky and Santi, 2008), selection among alternatives (Thompson-Schill et al., 2005), or executive functioning (Koechlin and Summerfield, 2007). Therefore, while no firm conclusions can be drawn from the absence of effects in the LIFG and LPTL, these null results do lend force to the increasing amount of data suggesting that the traditional neurophysiological model of language, revolving around Broca’s and Wernicke’s area, is neither necessary nor sufficient as a model for even the most basic linguistic processing.

Conclusion.

Surprisingly, direct investigations into the neural underpinnings of basic combinatorial processing in language have been virtually non-existent. The present paradigm introduces a powerful method for directly investigating these operations by allowing the linguistic expressions under consideration to be reduced to the absolute minimum – a simple adjective composed with a noun. Interestingly, our manipulation did not reveal any increases in activity within the more traditional “Broca’s” or “Wernicke’s” areas during basic combinatorial processing. Thus, our results add to the growing body of evidence that complete neurophysiological models of language must take into consideration a more extended network of neural regions than this simple bi-nodal configuration. Instead, our results indicate that both the LATL and vmPFC play a prominent role in basic linguistic composition. Activity within these regions was significantly greater during composition compared to matched controls, with increased LATL activity preceding increased vmPFC activity during combinatorial processing. As past work has implicated the LATL in syntactic composition and the vmPFC in
semantic composition, these results are consistent with the hypothesis that the LATL supports basic syntactic structure building while the vmPFC subserves the fundamental creation of linguistic meaning. Future work must now be aimed at building upon this foundation in order to construct a more solid and complete understanding of the neural mechanisms that underlie the comprehension and production of more complex linguistic expressions.
References


anterior temporal area analyzed by an fMRI-assisted MEG multidipole method. NeuroImage 44:1093-1102.


Wernicke C (1874) Der aphatische symptomencomplex: eine psychologische studie auf anatomischer basis.: Breslau: Cohn und Weigert.
Figure 1. Experimental design. Our design crossed Task (Composition vs. List) and Number of words (two vs. one). In each trial, participants indicated whether the target picture matched the preceding words. To satisfy this criterion, in the Composition task, *all* preceding words were required to match while in the List task, *any* matching word sufficed. A total of six colors and 25 shapes were randomly combined and used as stimuli. Half of the target pictures matched while half did not. Only activity recorded at the matched nouns (*boat*) was analyzed.
**Figure 2.** Behavioral results. Reaction time (A) and accuracy (B) data were submitted to a 2 x 2 repeated-measures ANOVA with Task (Composition vs. List) and Number of words (one vs. two) as factors. We observed a significant interaction between the two factors for reaction time (\(F(1,19) = 89.2\)), with post-hoc tests revealing slower responses in the two-word List condition (\(p < 0.0001\)) and faster responses in the two-word Composition condition (\(p < 0.001\)), compared to matched one-word controls. No significant effects were found for accuracy.
Figure 3. MEG sensor data. The average evoked response to the critical noun is displayed for each condition – (A) two-word Composition; (B) one-word Composition; (C) two-word List; (D) one-word List. Canonical visual response peaks and field patterns (M100 and M170 - Tarkiainen et al., 1999; Pylkkänen and Marantz, 2003) are visible at approximately 100ms and 150ms in all four conditions. Subsequent M250 and M350 peaks and field patterns (Pylkkänen et al., 2002; Pylkkänen and Marantz, 2003) are also present across all four conditions. A typical anterior midline field pattern (Pylkkänen and McElree, 2007) can be seen accompanying a sustained increase in activity from 300-500ms within the two-word Composition condition. This field pattern is absent from the other three conditions.
Figure 4

A

LATL

B

vmPFC

C

RATL

D

LIFG

E

LPTL

No Significant Clusters

No Significant Clusters

184-255ns

331-480ns

184-403ns
**Figure 4.** ROI results. Localized activity is shown for the five ROIs during the comprehension of the critical nouns, averaged across participants. Shaded regions denote significant clusters of combinatorial activity as identified by a cluster-based permutation test (Maris and Oostenveld, 2007) applied to the entire interval of 0-500ms. Within the left anterior temporal lobe (LATL) (A), a significant cluster of combinatorial activity was found from 184-255ms ($p = 0.039$), within which two-word Composition activity was significantly greater than one-word Composition activity ($p = 0.005$) and activity in the List conditions did not differ. A later cluster of combinatorial activity was found in the ventro-medial prefrontal cortex (vmPFC) (B) from 331-480ms ($p = 0.014$). In this time period, localized activity was again significantly greater for the two-word Composition condition compared to the one-word Composition condition ($p = 0.039$). No differences in activity were observed in the List task. Within the right anterior temporal lobe (RATL) (C), two significant clusters of time points were identified by the interaction permutation test (184-246ms, $p = 0.008$; 329-403ms, $p = 0.004$). Across the entire time window spanning these two clusters (184-403ms), activity in the two-word Composition condition was significantly greater than in the one-word condition ($p = 0.003$). However, unlike in the LATL and vmPFC, activity in this region also showed a significant difference between the two one-word conditions during the identified time window ($p = 0.041$), suggesting a task-related suppression of activity in the one-word Composition condition as opposed to an increase in activity during the two-word Composition condition. No significant clusters of combinatorial activity were found in either the left inferior frontal gyrus (LIFG) (D) or left posterior temporal lobe (LPTL) (E) ROIs.
Targeted tests within the Composition task alone also failed to find any periods of significantly greater activity in the two-word Composition condition compared to the one-word Composition condition in either region at any time.
Figure 5. Full brain analyses. Plotted regions denote the difference in average amplitude between two-word and one-word conditions for all space-time regions in which two-word activity was reliably greater than one-word activity ($p < 0.05$, uncorrected) for at least 10ms over 10 spatial neighbors. For clarity, non-cortical sources have been removed. Results within the Composition task (A) conform to our ROI analyses and reveal clear a left anterior temporal lobe (LATL) effect from 150-250ms and a clear ventro-medial prefrontal cortex (vmPFC) effect from 300-450ms. A sustained difference in activity can also be seen in the right anterior temporal lobe (RATL) from 100-400ms. Additionally, increased activity was also observed within the left superior parietal lobe throughout much of the time interval. Within the List task (B), relatively modest increases in two-word activity were seen in the right inferior frontal gyrus and left inferior temporal lobe from approximately 200-300ms.
Figure 6

A
Half 1
203ms

B
Half 2
230ms

C
Half 1
421ms

D
Half 2
412ms
Figure 6. Sensor analysis results. Each row corresponds to one test from the split-half cross-validation analysis. Waveforms represent the RMS of the sensors corresponding to the dominant field patterns present during the LATL (A, B) and vmPFC (C, D) source space effects. Shaded regions denote significant clusters of combinatorial activity as identified by a cluster-based permutation test (Maris and Oostenveld, 2007). The field pattern from the grand average of the two-word Composition condition is shown for each significant cluster. Within the left hemisphere sensors, the first data half (A) produced a significant cluster of combinatorial activity from 195-215ms ($p = 0.036$), within which two-word Composition activity was significantly greater than one-word Composition activity ($p = 0.038$) and activity in the List conditions did not differ. In the second data half (B), a significant cluster of combinatorial activity was identified from 220-241ms ($p = 0.013$) in which two-word Composition activity was significantly greater than one-word Composition activity ($p = 0.004$) and activity in the List conditions did not differ. Within the frontal sensors, the analysis of the first data half (C) revealed a cluster of combinatorial activity from 396-448ms ($p = 0.011$). In this time period, sensor activity was again significantly greater for the two-word Composition condition compared to the one-word Composition condition ($p = 0.046$). No differences in activity were observed in the List task. In the second data half analysis (D), a significant cluster of combinatorial activity was found from 388-442ms ($p = 0.050$). Again, two-word Composition activity was significantly greater than one-word Composition activity within this cluster ($p = 0.015$), while activity in the List conditions showed no significant differences.
Chapter 2: Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading (Revision submitted to Cerebral Cortex)

Abstract

Language is experienced primarily through one of two mediums – spoken words and written text. Though substantially different in form, these two linguistic vehicles possess similar powers of expression. Consequently, one goal for the cognitive neuroscience of language is to determine where, if anywhere, along the neural path from sensory stimulation to ultimate comprehension these two processing streams converge. In the present study, we investigate the relationship between basic combinatorial operations in both reading and listening. Using magnetoencephalography, we measured neural activity elicited by the comprehension of simple adjective-noun phrases (*red boat*) using the same linguistic materials and tasks in both modalities. The present paradigm deviates from previous cross-modality studies by investigating only basic combinatorial mechanisms – specifically, those evoked by the construction of simple adjective-noun phrases. Our results indicate that both modalities rely upon shared neural mechanisms localized to the left anterior temporal lobe (lATL) and left angular gyrus (lAG) during such processing. Further, we found that combinatorial mechanisms subserved by these regions are deployed in
the same temporal order within each modality, with IATL activity preceding IAG activity. Additionally, modality-specific combinatorial effects were identified during initial perceptual processing, suggesting top-down modulations of low-level mechanisms even during basic combination.
Introduction

Physically, linguistic expressions primarily impinge upon the human consciousness in one of two ways, through speech – consisting of vibrations in the air – or reading – conveyed through photons reflecting off of a physical surface. Despite these initial differences, both methods of expression ultimately give rise to similar mental states in the mind of a comprehender. While there have been many investigations into the convergence of these input streams during single word comprehension (e.g. Marinkovic et al. 2003) and several recent studies comparing the processing of complex linguistic expressions across modalities (e.g. Jobard et al. 2007), the extent to which basic combinatorial mechanisms in language are shared across modalities is unknown. Though past results generally support the view that speech and reading rely upon a common set of neural mechanisms, systematic discrepancies between neural effects generated by auditory and visual processing of complex expressions (e.g. Michael et al. 2001) make it unclear whether past similarities reflect shared basic combinatorial mechanisms or a common reliance on more complex processes such as working memory and cognitive control. Disentangling these possibilities is vital for determining the extent to which results acquired in one domain of language processing can be applied to the other; a question especially important within the fields of psycho- and neuro-linguistics as the majority of linguistic experiments have been carried out within the visual modality alone.

In the present work, we assess the extent to which basic linguistic combinatorial operations are shared between auditory and visual processing. Using a
minimal MEG paradigm, we measured neural activity associated with the comprehension of a simple noun (e.g. boat) and identified increases in this activity when the noun was combined with a preceding adjective (e.g. red) compared to when it was preceded by a non-word (e.g. xhl) or a non-compositional list control (e.g. cup). Within each subject, combinatorial processing was measured in both the auditory and visual modality using the same task and linguistic materials. Thus, we were able to search for evidence of shared combinatorial mechanisms by comparing between the two sets of results and identifying cortical sources that were preferentially active during basic linguistic combination regardless of modality.

*Sensory Level Operations.*

Clearly, given the vast differences in the physical nature of speech and written words, initial sensory processing of the two must diverge to a large degree. During reading, a visual word must first be broken down into basic visual units such as lines and intersections. These units must then be constructed into increasingly complex visual representations such as letters, graphemes and finally an orthographic representation of the word (see Dehaene 2009). Similarly, when listening to a spoken word, linguistic input must first be analyzed into basic auditory units such as different frequency bands. These features must then be combined into progressively larger and more complex segments such as phonemes, syllables, and, eventually, phonetic representations of words (see Hickok and Poeppel 2007). Unsurprisingly, these vast functional differences give rise to consistent neuroimaging differences between the two modalities. In general, though many regions are active during both visual and
auditory language processing, reading single words shows increased activity in bilateral occipital and inferior temporal regions compared to listening, which in turn evokes more activity bilaterally in the superior temporal gyri (e.g. Petersen et al. 1988; Booth et al. 2002, 2003). Temporally, there is strong evidence from electrophysiological studies that these differential regions are involved in early processing stages, generally completing around 200ms following the presentation of the stimulus.

MEG measures of estimated neural activity indicate that, prior to 200ms, visually presented words primarily evoke activity in the medial, posterior occipital cortex followed by the left inferior temporal cortex (see Salmelin 2007). The initial occipital activity peaks at approximately 100ms and has been shown to reflect the basic analysis of visual features as the amplitude of this neural measure increases with the noise present in the stimulus and the complexity of the presented objects (Tarkiainen et al. 1999; Tarkiainen et al. 2002). The second component localizes to the fusiform gyrus at approximately 150ms and responds preferentially to letter strings compared to symbols (Tarkiainen et al. 1999; Tarkiainen et al. 2002). This component is hypothesized to reflect the construction of complex visual representations from initial visual features, as this activity has also been shown to increase during the processing of complex non-linguistic stimuli such as faces (e.g. Xu et al. 2005) and can be modulated by the orthographic and even morphological properties of words (Zweig and Pylkkänen 2009; Solomyak and Marantz 2010).

During auditory speech comprehension, similar initial computations (i.e. the conversion of raw sensory data into primitive features followed by the construction of
complex perceptual representations) are also hypothesized to monopolize processing before 200ms, however, this activity is localized within the superior temporal cortex as opposed to the posterior occipital and inferior temporal regions (see Salmelin 2007). MEG studies have revealed an initial evoked component that responds to basic acoustic and phonetic properties of auditory words peaking at approximately 100ms (Kuriki and Murase 1989; Parviainen et al. 2005). This component has been localized to the superior temporal cortex bilaterally and is hypothesized to reflect the initial breakdown of complex auditory stimuli into basic perceptual features (Salmelin 2007). The timing and nature of secondary processing of acoustically complex stimuli has been inferred more indirectly than in the visual modality, as auditory studies have generally employed mismatch paradigms to investigate early acoustic processing. Results from these studies have found an early component that responds to acoustically unexpected stimuli occurring at approximately 180ms. This component has been shown to be sensitive to phonological categories (Phillips et al. 2000) and native compared to non-native vowels (Näätänen et al. 1997) and has also been localized to the superior temporal cortex bilaterally. The neural mechanisms supporting the creation of syllable and ultimately whole word phonological forms from these basic features have been difficult to pin down, however, hemodynamic investigations indicate that the superior temporal sulcus is integrally involved in such operations (Hickok and Poeppel 2004; Liebenthal et al. 2005). Thus, it is clear from both a theoretical and empirical standpoint that the neural processing streams evoked by visual and auditory linguistic items are highly divergent in the initial stages, both in terms of function and location. Thus, we expect to see the
clearest evidence for shared combinatorial mechanisms following this initial processing phase.

**Lexical Level Operations.**

Before combinatorial operations can proceed, however, perceptual representations are first connected to stored lexical meanings that then enter into various syntactic and semantic combinatory operations, ultimately leading to the comprehension of a full linguistic expression. Hemodynamic comparisons between auditory and visual lexical tasks and manipulations, such as semantic priming, repetition, relatedness judgments, and lexical decision, have in general found many shared regions of activity across modalities primarily in the left inferior frontal gyrus (LIFG) (Buckner et al. 2000; Booth et al. 2002; Price et al. 2003), left middle superior temporal cortex (Howard et al. 1992; Booth et al. 2002; Price et al. 2003; Sass et al. 2009) and left inferior temporal cortex (Buckner et al. 2000; Price et al. 2003; Cohen et al. 2004). In terms of timing, lexical access effects within these regions typically appear following early initial sensory processing (e.g. Marinkovic et al. 2003), and a wide variety of lexical tasks and manipulations have produced similar electrophysiological effects within both modalities. For example, word repetition (350ms – Rugg and Nieto-Vegas 1999), the processing of inflected words (400ms – Leinonen et al. 2009), semantic priming (400ms – Anderson and Holcomb 1995; Gomes et al. 1997), linguistic memory probes (560ms – Kayser et al. 2003), and the processing of semantically related and unrelated word lists (Vartiainen et al. 2009) have all evoked responses with similar topographies and timings across modalities.
These results have led to the predominant view that the initially divergent perceptual pathways converge upon a single mental lexicon (e.g. Caramazza et al. 1990; Price 2000; Marinkovic et al. 2003) – a position sometimes taken as assumed (e.g. Hagoort and Brown 2000b; Spitsyna et al. 2006). However, neurophysiological data also supports an alternate hypothesis in which there are two distinct semantic systems, one for each modality. Several patients have been identified who lack the ability to access various types of semantic knowledge when processing linguistic input in one modality but not the other (e.g. Warrington 1975; Beauvois 1982; Shallice 1987; Sheridan and Humphreys 1993; Marangolo et al. 2004). Because these deficits are not accompanied by perceptual impairments and they do not appear to depend upon the type of semantic information required, several researchers have posited the existence of two separate, modality-specific semantic systems to explain the pattern of results (e.g. Shallice 1988; Druks and Shallice 2000), though the necessity of this dichotomy has been questioned (e.g. Caramazza et al. 1990).

*Combinatorial Level Operations.*

While the debate regarding multiple versus unitary semantic systems remains unresolved, it should be made clear that the outcome does not force an answer one way or another to the present question of whether or not basic linguistic combinatorial operations are shared across modalities. Just as divergent perceptual processes might converge upon a shared lexicon, so too might divergent lexicons submit to operations by a single combinatorial engine. Conversely, it is logically possible, though somewhat unparsimonious, that different input modalities may utilize a shared lexicon
while still relying on divergent mechanisms to combine these representations. To reiterate Marr’s famous distinction (Marr 1982), even if the computational profiles of auditory and visual combinatorial mechanisms are *identical* – the same inputs go in and the same outputs come out – their implementations may be entirely distinct. For example, written words can connect to meaning either by way of a phonological code or orthography alone (Saffran and Marin 1977; Dérouesné and Beauvois 1979). Such dual routes between auditorily and visually based processing might reflect the different developmental timelines of reading and listening – between modalities, common electrophysiological measures evoked during combinatorial linguistic processing exhibit different properties during development (Holcomb et al. 1992) – or they might reflect the vastly different temporal structure of speech and writing. During speech comprehension, language enters the combinatorial system in the form of fluid auditory stimulation that unfolds continuously in time. The comprehender has little control over the speed or fluency of the input, and thus the combinatorial engine must be able to flexibly adapt on the fly to varying rates of informational input. In fact, even within the auditory modality alone, it has been suggested that multiple neural mechanisms are needed to parse the highly variable temporal properties of continuous speech (e.g. Poeppel 2003). Contrastingly, a comprehender encounters visual language in a static manner that allows for a higher degree of control over the rate that information is processed by the system.

These temporal differences have led past researchers (e.g. Hagoort and Brown 2000b) to suggest that different combinatorial processes operate within each modality. This hypothesis is bolstered by the finding that altering the structure of visual stimuli
to temporally reflect speech (i.e. dynamically unfolding over time) creates significantly more errors in comprehension than when it is presented in a more natural (i.e. static) format (Lee and Newman 2010). Also, semantic priming effects have been shown to increase with SOA during auditory processing and decrease with longer SOAs in visual processing (Anderson and Holcomb 1995), and temporal anaphora are more difficult to comprehend during speech compared to reading (Jakimik and Glenberg 1990). One possibility is that these effects reflect a deep neural divide in how the brain converts visual and auditory input from form to meaning. Perhaps the static nature of writing lends itself to neural mechanisms that can combine representations in parallel, while auditory processing must instead rely upon more serially arranged processes that reflect the temporally extended nature of speech. On the other hand, it is also possible that these effects stem instead from differences in complex, domain-general abilities such as working memory and cognitive control that operate in concert with a shared set of basic combinatorial mechanisms to process the variable input. At the present time, it is difficult to adjudicate between these two possibilities as no previous studies have directly investigated the extent to which basic combinatorial linguistic mechanisms are shared across modalities.

In general, the vast majority of psycho- and neuro-linguistic investigations contrast the processing of either complex or unexpected linguistic expressions (and often both) with simpler, but still structurally complex, controls. Thus, it is often difficult to confidently disentangle effects associated with basic combinatorial mechanisms from those resulting from more complex processing, such as compensation or repair strategies. Direct neurolinguistic comparisons between
auditory and visual combinatorial linguistic processing have overwhelmingly investigated either expectedness or complexity manipulations and in both cases canonical results from single modality investigations have been reliably replicated across modalities. For example, increased activity in the LIFG during the processing of object-relative compared to simple expressions was observed in fMRI for both modalities (Michael et al. 2001; Constable et al. 2004). Semantic priming effects in the LIFG and temporal cortex, previously reported in single word studies (e.g. Booth et al. 2002), were also observed following the repetition of sentences that overlapped in semantic content (Devauchelle et al. 2009). Within electrophysiology, both the N400 – canonically elicited by sentence final incongruous words (e.g. Kutas and Hillyard 1980) – and the P600 – consistently evoked following syntactically difficult or anomalous constructions (e.g. Osterhout et al. 1994) – have been observed in both modalities within the same set of subjects (N400 – Holcomb et al. 1992; Hagoort and Brown 2000b; Balconi and Pozzoli 2005) (P600 – Hagoort and Brown 2000a; Balconi and Pozzoli 2005).

While the convergent nature of these results certainly suggests that mechanisms recruited during the processing of complex linguistic expressions are shared across modalities, it is unclear exactly which aspect of such processing these results reflect. Specifically, it is unclear whether or not these results bear directly upon basic combinatorial processing. Both violation and complexity manipulations might produce results related to basic processing, however, there are a myriad of other processes that might differ between conditions in these manipulations as well, such as the amount of working memory, conflict resolution, or executive control required to
process the input. These complexities make it difficult to determine if observed commonalities arise from shared basic combinatorial mechanisms or a mutual dependence on more complex abilities. While most studies have highlighted effects shared across modalities, many investigations have also found differences during auditory and visual combinatorial processing both in the topography of electrophysiological effects (e.g. Niznikiewicz et al. 1997; Hagoort and Brown 2000b; Hagoort and Brown 2000a) and the pattern of hemodynamically activated voxels (Michael et al. 2001; Constable et al. 2004; Lindenberg and Scheef 2007; Buchweitz et al. 2009). Auditory effects generally evoke stronger (Holcomb et al. 1992; Hagoort and Brown 2000b; Michael et al. 2001; Spitsyna et al. 2006) more bilateral effects (Hagoort and Brown 2000b; Carpentier et al. 2001; Michael et al. 2001; Constable et al. 2004; Buchweitz et al. 2009) that have been observed to onset earlier and extend longer than their visual counterparts (Holcomb and Neville 1990, 1991; Osterhout and Holcomb 1993), leading several researchers to suggest that additional combinatorial mechanisms are invoked during auditory language comprehension compared to reading (e.g. Hagoort and Brown 2000b; Michael et al. 2001; Constable et al. 2004). On the other hand, several studies have also found that the processing of complex linguistic stimuli activates a wider region of the LIFG when presented visually (encompassing both BA 44 and 45) than auditorily (BA 45 only) (Carpentier et al. 2001; Constable et al. 2004). Thus, while past comparisons of auditory and visual combinatorial linguistic processing have found relatively consistent patterns of both shared and distinct neural activity, the complex nature of the manipulations used in these comparisons makes it difficult to assign specific functional roles to either.
In general, there has been a noticeable dearth of studies investigating basic combinatorial mechanisms (i.e. those employed during the comprehension of simple linguistic expressions), even within a single modality. In total, we could find only two cross-modal studies that treat simpler linguistic processing as the condition of interest as opposed to the baseline. Both studies contrasted the comprehension of sentences with unstructured lists of words or non-word stimuli, and both sets of results are consistent with the majority of unimodal manipulations of the same type (e.g. Friederici et al. 2000), finding increased activity in the bilateral anterior temporal lobes during sentence comprehension (Jobard et al. 2007; Lindenberg and Scheef 2007). Compared to manipulations of complexity or expectedness, these results more directly suggest that basic combinatorial linguistic mechanisms are shared across modalities, however, this evidence is still somewhat unclear as the level of complexity of the processing investigated in these studies (i.e. the comprehension of complete sentences for both and stories in one case) potentially involves a large number of diverse operations. Thus, at the present time, while there is a fair amount of evidence suggesting that certain facets of linguistic combinatorial processing may draw upon a common set of neural circuits in both visual and auditory processing, it is currently unknown whether basic combinatorial mechanisms (i.e. those responsible for the construction of minimal linguistic phrases) are shared across modalities.

The Paradigm.

The present paradigm was developed precisely in order to target such basic mechanisms (see Bemis and Pylkkänen 2011 and Figure 1). In our design, the neural
activity of interest is that evoked by a simple, object denoting noun (e.g. boat), and the fundamental comparison is between the presentation of this noun within a minimal compositional environment (i.e. preceded by a simple color adjective) compared to a non-compositional control (i.e. preceded by either a non-pronounceable consonant string in the visual modality or a burst of noise in the auditory modality). In order to ensure that subjects remained attentive to the stimuli, we asked them to judge whether or not a colored picture following the noun matched the linguistic stimuli that preceded it. In the adjective-noun condition, both the color and shape were required to match while in the one-word condition only the shape was relevant to the decision. To control for the difference in lexical-semantic material in these two conditions, we also had subjects complete a control task in which the adjectives were replaced by length-matched nouns. Subjects were then instructed to determine whether or not the shape of the following picture was denoted by any of the words that preceded it. Each stimulus was presented sequentially allowing us to record neural activity evoked at the point of basic combination (i.e. at the nouns following an adjective) compared to physically identical stimuli for which no combination occurred (i.e. nouns following a noun or non-word).

To compare combinatorial processing between modalities, we administered this paradigm twice to each subject, once with auditory linguistic stimuli and once with visual linguistic stimuli, while the tasks and target pictures were held constant. If basic combinatorial mechanisms are shared across modalities, we expect to find cortical regions for which localized MEG activity is greater during the combinatorial condition compared to the control conditions, regardless of the modality of the
linguistic stimuli. If, on the other hand, previously identified shared effects reflect more complex processes, then we might expect to see previously observed modality-specific effects augmented in the present comparison.
Materials and Methods

Participants.

Twenty-two non-colorblind, native English speakers participated in the study (13 female, average age: 25.6). All had normal or corrected-to-normal vision and gave informed consent.

Visual Stimuli.

During the visual tasks (see Figure 1), each trial contained a small fixation cross, an initial word or non-word, a critical noun, and a target shape. The initial word or non-word varied by condition and was either an adjective (two-word Composition condition), noun (two-word List condition), or unpronounceable consonant string (one-word conditions). Twenty one-syllable nouns were employed in the second (critical) stimulus position (disc, plane, bag, lock, cane, hand, key, shoe, bone, bell, boat, cross, cup, flag, fork, heart, lamp, leaf, star, house). Adjectives in the two-word Composition condition – (red, blue, pink, black, green, brown) – were matched in length with the nouns used as first stimuli in the List task – (cane, lamp, disc, fork, bone, heart). In both tasks, one-word trials were produced by substituting each adjective or initial noun with a corresponding unpronounceable consonant string of the same length – (xkq, qxsw, mtpv, rjdnw, wvcnz, zbxly). The sets of adjectives, non-words, and first-stimuli List nouns were each created in 28-pt non-proportional Courier font and were all manually sized to cover the same amount of filled area on the screen. Target shapes were manually created to depict canonical, unambiguous
representations of the objects denoted by the nouns and were filled in with one of the six colors denoted by the adjectives. Three versions of each target were then created by applying a random scaling factor between 105% and 115% and a random rotation of 0°-360° to the original figure. All stimuli were presented using Psychtoolbox (http://psychtoolbox.org/; Brainard 1997; Pelli 1997) and were projected approximately 50cm from the subject’s eye. Words subtended between 2° and 4° while target shapes were larger, subtending between 6° and 10°.

During each task, subjects viewed 160 trials, 80 of each trial type. All conditions contained an equal number of trials in which the target shape matched or did not match the preceding words while the two-word Composition condition additionally divided the non-matching trials equally among those that did not match the adjective and those that did not match the noun. In these trials, all target shapes matched at least one of the preceding words. During each condition, each of the 20 critical nouns was used four times, twice in matching and twice in non-matching trials. Trial and stimuli lists were randomized and constructed separately for each subject.

*Auditory Stimuli.*

The auditory and visual paradigms had identical trial and task structures, and the same stimuli lists were used for both presentations within a subject’s session. The only difference in the auditory paradigms was that the first two stimuli were converted into spoken words (see Figure 1). Each of the twenty nouns and six adjectives used in the visual tasks was recorded by a female speaker using neutral intonation. The average length of the twenty critical nouns was 569ms (56ms std.). The average length of the
six adjectives was 540ms (28ms std.). The six nouns used in the auditory List task remained the same as those used in the visual version (538ms [36ms std.]). These two sets of stimuli (the adjectives and initial nouns) were matched in length ($p > 0.60$, two-tailed paired t-test). Non-pronounceable consonant strings were replaced by pink noise. Pink noise is characterized by a falloff in spectral power inversely proportional to the frequency. This is in contrast to white noise, for which the power is constant across all frequencies, and red (or Brownian) noise for which the power decrease is inversely proportional to the square of the frequency. An initial five-second recording of pink noise (from http://www.mediacollege.com/audio/noise/download.html) was split into six distinct segments matched in duration to the adjectives.

Procedure.

During the experiment, subjects performed four separate blocks of trials, two in each modality and two of each task. Blocks of the same task were always presented sequentially and modality order within each task was constant for each subject. Overall order of tasks and modality order within tasks was counterbalanced across all subjects. Thus, there were four possible orderings of the four block types. Before the experiment, subjects practiced their first task in their first modality outside of the MEG room. Though subjects were made aware of the existence of a second task and modality at this time, no specific instructions regarding the second task were given before the completion of the first task. Instructions and practice for this second task were then given following the completion of both blocks of the first task, while subjects remained in the machine. Prior to recording, subjects’ head shapes were
digitized using a Polhemus Fastrak 3D digitizer (Polhemus, VT, USA). The digitized head shape was then used to constrain source localization during analysis by co-registering five coils located around the face with respect to the MEG sensors. Additionally, electrodes were attached 1 cm to the right of and 1 cm beneath the middle of the right eye in order to record the vertical and horizontal electrooculogram (EOG) and detect blinks. Both electrodes were referenced to the left mastoid.

MEG data were collected using a using a whole-head 157-channel axial gradiometer system (Kanazawa Institute of Technology, Tokyo, Japan) sampling at 1000Hz with a low-pass filter at 200Hz and a notch filter at 60Hz. All visual stimuli besides the target shapes were presented for 300ms and were each followed by a 300ms blank screen. Corresponding initial auditory stimuli (i.e. noise, adjectives, and list nouns) were presented for 600ms while critical nouns were presented for 700ms. These durations did not vary with the length of the individual stimuli and were followed in both cases by 300ms of silence. In both modalities, target shapes appeared at the end of each trial and remained onscreen until the subject made a decision. Subsequent trials began after a blank screen was shown for a variable amount of time. This delay followed a normal distribution with a mean of 500ms and a standard deviation of 100ms (see Figure 1). The entire recording session lasted approximately 1 hour.

Data Acquisition.

MEG data from the 100ms prior to the onset of each critical noun to 600ms post onset were segmented out for each trial. Raw data were first cleaned of potential artifacts by rejecting trials for which the subject answered either incorrectly or too slowly (defined
as more than 2.5 seconds after the appearance of the target shape) or for which the maximum amplitude exceeded 3000fT or for which the subject blinked within the critical time window, as determined by manual inspection of the EOG recordings. Remaining data were then averaged for each subject for each condition and band-pass filtered between 1 and 40Hz. For inclusion in further analysis, we required that subjects show a qualitatively canonical profile of evoked responses during the processing of the critical items. This profile was defined as the appearance of robust and prominent initial sensory responses. In the visual modality, we required either the M100 or M170 field pattern (e.g. Tarkiainen et al. 1999; Pylkkänen and Marantz 2003) to be present in the time window of 100-200ms following the critical stimuli. In the auditory modality, subject data were required to show a prominent bilateral M100 response situated over the primary auditory cortex, preceded by a similar component of opposite polarity (e.g. Poeppel et al. 1996). In order to assess this criterion, preliminary grand average waveforms were constructed for each subject by averaging over all conditions. Three subjects failed to meet this requirement and were excluded from further analysis.

Minimum Norm Estimates.

Distributed minimum norm source estimates served as our primary dependent measure. After preprocessing, a source estimate was constructed for each condition average using L2 minimum norm estimates calculated in BESA 5.1 (MEGIS Software GmbH, Munich, Germany). The channel noise covariance matrix for each estimate was based upon the 100ms prior to the onset of the noun in each condition average.
Each minimum norm estimate was based on the activity of 1426 regional sources evenly distributed in two shells 10% and 30% below a smoothed standard brain surface. Regional sources in MEG can be regarded as sources with two single dipoles at the same location but with orthogonal orientations. The total activity of each regional source was then computed as the root mean square of the source activities of its two components. Pairs of dipoles at each location were first averaged and then the larger value from each source pair was chosen, creating 713 non-directional sources for which activation could be compared across subjects and conditions. Minimum norm images were depth weighted as well as spatio-temporally weighted using a signal subspace correlation measure (Mosher and Leahy 1998).

Data Analysis.

We determined effects related to basic combinatory processing using a two-stage analysis. Because the timing of apparently functionally equivalent effects can vary between modalities (e.g. Holcomb and Neville 1990; Osterhout and Holcomb 1993; Marinkovic et al. 2003), to identify shared combinational neural activity we first performed a full-brain analysis over activity from each cortical source within each modality separately and then collapsed the results over time. This process produced a full-brain map for each modality that identified which sources were preferentially active during combination (i.e. which sources showed significantly more activity during the two-word condition during the composition task and no difference during the list task), regardless of the time at which the combinational activity occurred. We then merged the results of these two analyses into a single map, assigning each cortical
source one of four possible designations: combinatorially important regardless of modality, combinatorially important for reading only, combinatorially important for speech only, or not combinatorially important. Next, we investigated the time-course of these effects by dividing up the resulting map into separate regions of interest and performing a cluster-based permutation test (Maris and Oostenveld 2007) within each region for each modality. This test produces clusters of time points for which the activity localized to a given region was significantly greater during composition compared to the non-combinatorial controls. Thus, our two-step analysis allowed us to identify sources of combinatorial activity that were shared between modalities even if their associated effects occurred at different times for each modality. The follow-up ROI analyses then allowed us to determine the temporal profile within each modality of the identified combinatorial effects.

For the initial full-brain comparison, we first performed a paired t-test within each task at every time point during the critical interval for every cortical source. We then classified sources as significantly active during a task if there was a cluster of at least 20ms for which the p-value of each t-test was less than 0.05 and the two-word condition had more activity than the one-word condition. Thus, based upon this criterion, each source was initially classified as either significantly active or not for each of the four tasks. Within each modality, a source was then only considered preferentially active during combinatorial processing if it was significantly active in the composition task and not the list task for that modality. Finally, the resulting full-brain maps for each modality were combined to produce a single map that assigned each source a status as (1) preferentially active during combinatorial processing
regardless of modality, (2) preferentially active during visual combinatorial processing only, (3) preferentially active during auditory combinatorial processing only, or (4) not preferentially active during combination.

Note that the initial classification of each source – as preferentially activity during combination in either modality – was arrived at without directly comparing the activities in the two two-word tasks to each other. In other words, at no point is the red boat condition compared directly to the cup, boat condition in our analysis; instead each two-word condition is first contrasted with the paired one-word control and the resulting differences are then compared in order to designate each source as combinatorial or not. The reason for this approach, as opposed to performing direct comparisons between the two two-word conditions, arises from the blocked nature of our design. The purpose of the list task is to ascertain if increased activity identified during the two-word composition condition is actually due to combinatorial processing or merely the processing of two lexical items during a single trial. Thus, it is constructed to be maximally parallel to the composition task in this latter respect. However, this parallelism comes at the cost of requiring the subjects to perform a different task. Therefore, it is possible that the demands of this alternate task itself might introduce additional activity into the critical time period, thus confounding any direct intertask comparisons, such as between the red boat and cup, boat conditions. To mitigate this potential confound, we first contrast pairs of conditions within each task and only afterwards do we compare combinatorial to list processing. Under this approach, activity due simply to the processing of multiple lexical items will still emerge in both the composition and list comparisons, as any increases in two-word
over one-word activity will be carried forward to the next stage of analysis, however increases in activity due solely to the demands of one task or the other will be neutralized before the critical between-task comparison. Thus, while the list task serves as the control for the composition task, it is the task as a whole that serves as the control and not the two-word list condition by itself.

In the second analysis stage, we constructed regions of interest from the resulting combined full-brain alaysis by extracting all clusters of combinatorial sources that spanned at least 10 adjacent cortical locations. The time-course of activity localized to each region was then investigated within each modality using the same cluster-based permutation test as in our previous work (see Bemis and Pylkkänen 2011 for details). Briefly, this test analyzes the interaction of activity from the four conditions in order to identify time intervals for which there was significant combinatorial activity in a given region, corrected for multiple comparisons across the entire epoch. Additionally, we performed simpler follow-up cluster tests within each task to support these analyses (see Bemis and Pylkkänen 2011 for details). As the regions of interest had already been identified as combinatorial during the first phase of the analysis, the primary purpose of these ROI analyses was to characterize and compare the timing and strength of the combinatorial effects within each modality.
Results

Behavioral.

Reaction time and accuracy data were subjected to a 2x2x2 repeated-measures ANOVA with Modality (Visual vs. Auditory), Task (Composition vs. List), and Words (One vs. Two) as factors (see Figure 2). The only significant effect for accuracy was a main effect of modality ($F(1,18) = 8.753, p = 0.008$). Subjects were more accurate in the auditory modality compared to the visual modality (97.6% avg. [0.85% std.] vs. 95.4% avg. [3.39% std.]). Though no other effects were significant, it should be noted that accuracy data were qualitatively different between modalities. Within the Composition task accuracy did not differ significantly within the visual modality, though responses tended to be more accurate for the two-word condition than the one-word condition ($p = 0.37; 96.7\% \text{ avg.} [4.5\% \text{ std.}] \text{ vs. } 94.1\% \text{ avg.} [10.7\% \text{ std.}]$). For the auditory modality, on the other hand, subjects were significantly less accurate for the two-word Composition trials compared to the one-word condition ($p = 0.001; 97.0\% \text{ avg.} [2.0\% \text{ std.}] \text{ vs. } 98.7\% \text{ avg.} [1.3\% \text{ std.}]$). Within the List task, accuracy results were virtually identical across modalities. Subjects were more accurate for one-word trials in both modalities (Visual: $p = 0.017, 94.4\% \text{ avg.} [4.1\% \text{ std.}] \text{ vs. } 96.1\% \text{ avg.} [3.2\% \text{ std.}];$ Auditory: $p < 0.001, 96.2\% \text{ avg.} [2.0\% \text{ std.}] \text{ vs. } 98.4\% \text{ avg.} [1.4\% \text{ std.}]$). The non-significance of these effects within the larger ANOVA is likely due to the subjects being close to ceiling on all tasks.

Unlike accuracy, reaction times were qualitatively similar across modalities. For reaction times, we found a significant interaction between Task and Number of
words \((F(1,18) = 43.47, p < 0.001)\). This interaction was driven by a significant
decrease in response times for two-word Composition trials relative to one-word
Composition trials \((p = 0.015; 601\text{ms avg. [118ms std.] vs. 631\text{ms avg. [111ms std.]}\))
and a significant increase in response times for two-word List trials relative to one-
word List trials \((p < 0.0001; 732\text{ms avg. [153ms std.] vs. 651\text{ms avg. [130ms std.]}\))
Within the Composition task, the increase in reaction time for one-word trials was
significant in the visual modality \((p = 0.018; 563\text{ms avg. [103ms std.] vs. 612\text{ms avg.}
[117\text{ms std.]})\) but not the auditory modality \((p = 0.43; 639\text{ms avg. [124ms std.] vs.}
649\text{ms avg. [103ms std.]})\). In both modalities, reaction times were significantly slower
for two-word List responses compared to one-word List decisions (Visual: \(p < 0.001;\)
730\text{ms avg. [153ms std.] vs. 648\text{ms avg. [136ms std.]}\); Auditory: \(p < 0.001; 734\text{ms}
avg. [157\text{ms std.}] vs. 653\text{ms avg. [127\text{ms std.}]})
Overall, behavioral results in the visual modality replicated those found in our
previous experiment (Bemis and Pylkkänen 2011). In the visual List task, two-word
trials were harder than one-word trials, producing both slower and less accurate
responses. As before, this pattern was reversed within the Composition task, with two-
word trials proving easier than one-word trials. In the auditory modality, results were
essentially the same for the List task with two-word decisions coming both
significantly slower and less accurately than one-word responses, however, within the
Composition task the auditory responses showed no clear difference in difficulty
between the two conditions. Two-word trials were responded to slightly faster than
one-word trials, but subjects were significantly less accurate in doing so. The cause of
this difference across modalities within the Composition task is unclear. One
possibility is that this discrepancy reflects a general slowdown in auditory processing relative to visual processing (e.g. Marinkovic et al. 2003) such that the advantage of combination seen for the visual modality might not be available as readily within the auditory modality.

*General assessment of MEG data.*

Figure 3 depicts a qualitative overview of the MEG responses evoked during the critical items in each modality, averaged across all subjects. In the visual modality, canonical responses can be seen at ~100ms and ~150ms following the presentation of the critical word. These responses have been identified following the presentation of most complex visual stimuli and are thought to reflect early sensory and secondary processing of the visual word form (Tarkiainen et al. 1999; Pylkkänen and Marantz 2003). As can be seen, cortical activity during this time is localized to the primary visual cortex and inferior temporal-parietal cortex. These responses are then followed by the characteristic M250 and M350 patterns that are consistently evoked by the presentation of visual words (Embick et al. 2001; Pylkkänen et al. 2002). As expected (Marinkovic et al. 2003), activity during this later time window localizes primarily to the left temporal cortex.

In the auditory modality, characteristic early field patterns can be observed at ~50ms and ~100ms following the onset of the spoken word. As in previous work (e.g. Poeppel et al. 1996), we see that these early auditory responses are very similar in spatial extent and location but have opposite polarities. Activity during this time period localizes to the primary auditory cortex. As in the visual modality, activity then
spreads across the temporal lobe localizing to similar but slightly larger cortical regions. Thus, across both modalities the general pattern of activity very closely mirrors that observed in previous MEG comparisons of visual and auditory word processing (e.g. Marinkovic et al. 2003). Evoked activity initially localizes to the corresponding primary and associative cortices for each modality and then converges, largely within the temporal lobe, following these early responses.

Full-Brain Combinatorial Analysis.

The results of our full-brain analysis are shown in Figure 4A. We identified two cortical regions that reflect combinatorial activity in both modalities – one in the left anterior temporal lobe (lATL) and one in the left angular gyrus (lAG). Additionally, we observed a large region in the right posterior temporal and occipital cortices that reflected combinatorial activity in the visual modality alone. Auditory specific combinatorial sources were located anterior to the shared lAG region, near the primary auditory cortex. Additionally, a separate auditory-specific source abutting the inferior edge of the lAG region was identified. For simplicity, this latter source was incorporated into the larger lAG ROI in the subsequent temporal analysis.

Left Anterior Temporal Lobe ROI.

Our interaction cluster test identified a significant time region of combinatorial activity within the lATL (see Figure 4B) for the visual modality from 193-241ms ($p = 0.018$) and in the auditory modality from 267-357ms ($p = 0.049$). Follow-up permutation tests within each task supported these results and identified a significant
cluster of combinatorial activity in the visual composition task (191-299ms, \( p = 0.006 \)) and a marginally significant cluster of combinatorial activity within the auditory composition task (268-323ms, \( p = 0.086 \)). Neither test within the list task revealed any time windows for which activity during the two-word condition was substantially greater than the one-word condition (visual: all clusters with \( p > 0.25 \); auditory: all clusters with \( p > 0.8 \)).

A visual inspection of activity in the IATL, however, does appear to show an increase in two-word activity during the visual list task in a similar time window as the combinatorial effect identified above. Though this effect does not approach significance in any of our multiple-comparison tests, a post-hoc analysis directed specifically at this task and time interval does show a marginally significant increase in activity during the two-word list condition compared to the one word control (one-tailed \( t \)-test, 200-300ms, \( p = 0.06 \)). This increase might reflect the fact that, along with being implicated in combinatorial processing, the IATL has also been linked to the semantic processing of single words (e.g. Devlin et al. 2000; Rogers et al. 2006; Visser et al. 2010). Thus, it may be the case that this cortical region performs multiple functions during the comprehension of language, both associated with and orthogonal to combinatorial processing, and that the interaction of these multiple processes results in a combinatorial effect being overlaid upon more muted lexical-semantic process. Future research will hopefully be able to disentangle these potentially different processes and move towards a reconciliation of the, as of now, disparate literatures.

Overall, our analysis of IATL activity reveals that combinatorial processing preferentially drove activity within this region in both modalities. This effect was both
earlier and stronger during reading than when listening to speech. An inspection of the waveforms shows that the auditory lATL combinatorial activity peaks approximately 100ms after the visual activity and is more temporally extended than its visual counterpart. Qualitatively, these results are similar to many past results that find auditory effects to be both later and longer lasting than corresponding visual effects (e.g. Marinkovic et al. 2003).

**Angular Gyrus ROI.**
Within the lAG ROI (see Figure 4B), the interaction test revealed a cluster of combinatorial activity in the visual tasks that was marginally significant from 356-389ms ($p = 0.074$). A follow-up test within the composition task alone confirmed this effect, showing a similar cluster of combinatorial activity from 336-390ms ($p = 0.082$). No such cluster was observed in the list task (all clusters $p > 0.65$). In the auditory modality, a significant cluster of combinatorial activity was identified from 537-591ms ($p = 0.040$). (Additionally, a marginally significant cluster was identified from 283-326ms, $p = 0.052$, however, as this activity was contiguous both in time and space with an effect found in the auditory specific sources, it is discussed below in that context.) The follow-up test within the composition task alone did not reach significance due to a 9ms gap in the significant cluster identified by the interaction test for which the pairwise comparisons fell above the established cluster threshold. However, a targeted comparison between activity in the two compositions conditions at the time identified by the interaction test did show significantly greater activity in the two-word auditory composition condition compared to the one-word control ($p =
0.05, paired $t$-test). Within the list task, no clusters of increased two-word activity were identified by a follow-up cluster test (all clusters $p > 0.65$) and no significant increases were found for any 50ms time window between 400 and 600ms (tested at 10ms intervals, all tests $p > 0.20$). Thus, our results indicate the involvement of the lAG in combinatorial processing following lATL activity in both modalities. In the visual modality, combinatorial lAG activity followed lATL activity by approximately 150ms, while in the auditory modality the delay was slightly longer, ~250ms. As with the lATL, this general activity profile, with auditory effects both slower and more elongated than visual effects, matches previous findings (e.g. Marinkovic et al. 2003).

*Visual Specific Combinatorial ROI.*

Our initial full-brain analysis identified a large region of activity in the right occipital-temporal cortex that was preferentially active during the visual composition task. A cluster analysis within this region (see Figure 4B) revealed significant combinatorial activity in the visual modality from 141-189ms ($p = 0.011$). A similar cluster was also identified within the visual composition task alone, 134-202ms ($p = 0.022$). No such cluster was identified within the visual list task (all clusters with $p > 0.80$). Within the auditory modality, no time periods showed any type of effect in this region for any of the cluster tests (all clusters $p > 0.35$). While the observed effect extends across a relatively large section of cortex, the perceptual analysis of visual features has been associated with similar spatial extents in both fMRI (e.g. Grill-Spector et al. 2001) and MEG (e.g. Solomyak and Marantz 2010). Thus, these results indicate that the combinatorial effects in this region were restricted to the visual modality and occurred
approximately in both the location and time window most often associated with the perceptual parsing of complex visual objects (e.g. Tarkiainen et al. 1999; Solomyak and Marantz 2009).

**Auditory Specific Combinatorial ROI.**

The region identified by the full-brain analysis as active only during auditory combination fell close to the primary auditory cortex and was adjacent to the shared lAG ROI (see Figure 4B). A cluster analysis of this region within the auditory modality revealed a significant cluster of combinatorial activity from 287-401ms ($p = 0.039$) that was again confirmed by a follow-up test within the composition task alone (286-419ms, $p = 0.042$). No clusters of increased two-word activity were identified in the auditory list task. Within the visual modality, no significant clusters of combinatorial activity were identified by either the interaction or modality specific permutation tests (all clusters $p > 0.15$). As noted above, the effect identified in this region during listening temporally coincides with a marginally significant combinatorial lAG effect. We believe that this latter activity is an extension of the effect identified in this auditory-specific ROI as it occurs in sources adjacent to the auditory-specific ROI but is shorter in time (114ms vs. 43ms) and smaller in strength ($p = 0.04$ vs. $p = 0.05$). It is possible that the lAG activity during this time reflects a separate combinatorial mechanism that happens to be active at the same time in nearly the same place, however, the spatial detail of MEG is not sufficient to fully determine between these two possibilities. We believe that the present data is most consistent with one auditory combinatorial effect from approximately 300-400ms, centered near
the primary auditory cortex but potentially spreading into the lAG ROI, and another later combinatorial effect from approximately 550-600ms, centered in the lAG.
Discussion

The present study constitutes the first direct comparison of the neural correlates of basic linguistic combinatorial processing in both reading and listening. While there have been previous investigations into shared combinatorial mechanisms in language comprehension (e.g. Michael et al. 2001; Constable et al. 2004), none have reduced the scope of the studied processes to that of a simple adjective-noun phrase. In the present paradigm, we used MEG to record neural activity as subjects either listened to or read common, object-denoting nouns (boat) and measured increases in this activity during minimal compositional contexts (red boat) compared to non-compositional controls (xhl boat). To account for effects due solely to differences in lexical-semantic material during the combinatorial condition, we also compared a non-combinatorial list condition (cup, boat) to the same non-compositional control. In both modalities, we identified increased activity in the lATL and lAG associated with basic linguistic combination. A subsequent targeted ROI analysis within each modality revealed a temporal consistency to these effects such that combinatorial lATL activity preceded combinatorial lAG activity in both modalities. Between modalities, differential effects were confined to primary and associative sensory cortical regions and occurred prior to modality-independent combinatorial activity.

Thus, our results suggest that both auditory and visual processing of minimal linguistic phrases relies upon a shared set of basic combinatorial neural mechanisms that localize to the lATL and lAG, and that shared combinatorial effects previously identified in these regions during more complex processing (e.g. Humphries et al.
2001; Vandenberghe et al. 2002) reflect basic combinatorial mechanisms. Additionally, early perceptual processing appears to be modulated during the comprehension of basic linguistic phrases in both modalities.

*Modality-Independent Combinatorial Effects.*

Our initial full-brain analysis identified two neural regions – the lATL and lAG – that were significantly more active during composition compared to non-compositional controls in both the visual and auditory modalities. Targeted ROI analyses within both regions revealed clusters of time points for which activity elicited by the critical noun was significantly greater during composition than the non-combinatorial control. This result held for both modalities, and in both modalities increased lATL activity preceded increased lAG activity during basic combinatorial processing. Importantly, no significant increases were identified for activity within these regions during the non-compositional contrast at any point in either modality. In the visual modality, the increased lATL activity peaked at approximately 225ms (with a significant cluster of combinatorial activity occurring from 193-241ms) followed by increased lAG activity at approximately 375ms (356-389ms cluster). In the auditory modality, the combinatorial lATL effect occurred at approximately 300ms (267-357ms cluster) followed by increased lAG activity at ~560ms (537-591ms cluster). One likely explanation for the delay in auditory processing is that during listening a comprehender must wait for a word to unfold in time before it can be uniquely identified relative to its phonological competitors whereas in reading the entire word is immediately present. Typical models of auditory word recognition posit that higher-
level processing must wait until this uniqueness point has been reached (e.g. Marslen-Wilson 1987), and the temporal delay associated with waiting for the uniqueness point has been shown to affect the timing of later electrophysiological components such as the N400 (O'Rourke and Holcomb 2002). Thus, the temporal discrepancy we observe here might reflect the extra time needed during listening to accrue enough auditory signal to uniquely identify the incoming word – a suggestion offered previously to explain auditory effects found to be delayed relative to their visual counterparts (e.g. Marinkovic et al. 2003; Leinonen et al. 2009). Thus, our results suggest that visual and auditory language processing utilize shared combinatorial mechanisms during the comprehension of simple linguistic phrases and that these shared operations are subserved by the lATL and lAG.

Previous work has implicated both of these regions in combinatorial linguistic processing. The lATL has consistently been shown to be more activity during the processing of sentences compared to word lists (Mazoyer et al. 1993; Bottini et al. 1994; Stowe et al. 1998; Friederici et al. 2000; Humphries et al. 2001; Vandenberghhe et al. 2002) and was found to be significantly involved during basic composition in a previous, visual-only study using the present paradigm (Bemis and Pylkkänen 2011). Thus, the current results bolster the hypothesis that effects found in the lATL during sentential processing reflect basic combinatorial operations. Furthermore, they suggest that the neural mechanisms underlying these processes are shared between the auditory and visual modalities, a finding consistent with previous results derived from more complex expressions (e.g. Jobard et al. 2007; Lindenberg and Scheef 2007).
The lAG has also been identified as more involved during the comprehension of sentences compared to word lists (Bottini et al. 1994; Humphries et al. 2001; Humphries et al. 2005; Humphries et al. 2007), and, as with the lATL, was found to be preferentially active during basic composition in our previous study, though this effect was relatively weak (p = 0.11, Bemis and Pylkkänen 2011). Additionally, neurophysiological studies indicate that lesions in the lAG are associated with difficulties in processing complex sentences (Dronkers et al. 2004), leading to the suggestion that this region may play a role in integrating semantic information into context (Lau et al. 2008). Thus, the present results suggest that previous sentential-level effects observed in the lAG also reflect basic combinatorial operations and that these processes operate independent of the modality of input.

Beyond the shared spatial nature of these two effects, the consistency in their temporal ordering – lATL activity preceded lAG activity in both modalities – further supports the conclusion that these effects reflect shared combinatorial operations. While the gross nature of our contrast makes it difficult to parcel out different types of combinatorial operations to one region or the other, many existing models of sentence comprehension posit a two-step process in which syntactic composition precedes semantic combinatorial operations (e.g. Friederici 2002). The lATL has previously been linked to syntactic processing, with increased activity in this region corresponding to measures of syntactic complexity during natural story comprehension (Brennan et al. 2012) and decreased activity resulting from contexts that elicit syntactic priming (Noppeney and Price 2004). Contrastingly, the lAG has been found to be significantly more active during the comprehension of semantically
congruent sentences compared to syntactically well-formed but incoherent sentences (Humphries et al. 2007), and has thus been suggested to be critically involved in the integration of semantic information during language processing (Lau et al. 2008). Following these findings, the present results are consistent with a delineation of syntactic combinatorial processing to the lATL and semantic composition to the lAG. Of course, this suggestion must be extremely tentative at the present time, as the nature of the present manipulation does not allow for any direct insight into the partitioning of distinct combinatorial functions to distinct cortical regions.

It should be noted that while our previous study (Bemis and Pylkänen 2011) also identified a two-phase pattern of combinatorial activity in which the first phase consisted of increased lATL activity, the location of the most salient subsequent combinatorial effect was in the ventro-medial prefrontal cortex (vmPFC) and not the lAG. A targeted post hoc analysis of vmPFC activity in the present results, based on the spatial-temporal ROI identified in the previous study, did reveal a significant increase in activity during the two-word condition compared to the one-word control within the composition task (paired t-test, 300-400ms, $p < 0.05$). However, in the present study, this increase was accompanied by a corresponding increase in the list task as well (paired t-test, 300-400ms, $p < 0.01$). Thus, the strong interaction observed in the previous study was not replicated in the present results. Within the auditory domain, vmPFC activity exhibited the pattern predicted for combinatorial activity (i.e. activity in the two-word composition condition exceeded that in the one-word control to a greater degree than in the list task), however, results from a targeted post-hoc analysis of this activity did not reach significance (300-400ms, repeated-measures
ANOVA, all $F_s < 2$; paired $t$-tests within tasks, both $t < 1$). Thus, though our present results are not inconsistent with our past findings, they much less clearly indicate combinatorial processing in the vmPFC, as the significant increase observed in vmPFC activity during the composition task was also accompanied by a corresponding increase in the list task as well.

There are several possible explanations for this unexpected result. For one, the functional theory underlying the list task is much less clear than for the composition task, and so unexpected increases in activity might occur during the former depending upon how subjects perform this task. Perhaps a more clearly articulated functional theory of the list task in conjunction with an analysis of different behavioral metrics might help to explain the differences observed in vmPFC activity. At the present time, however, the exact form of such an explanation is unclear. Another possibility is that the increased length of the experiment (nearly twice as long as the original) as well as the interleaving of multiple modalities and tasks might have contributed to modulations of activity within the vmPFC, as task switching has been shown to affect activity in this region (e.g. Dreher et al. 2002) as well as nearby cortical and subcortical structures (e.g. Swainson et al. 2003; Crone et al. 2006). Unfortunately, a post-hoc analysis breaking down different task and modality orders lacked sufficient power to identify any significant effects that might support this hypothesis. Another deviation from the previous experiment was that in the present analysis we used a more conservative artifact rejection protocol, both in terms of a stricter amplitude rejection criterion and the use of ocular EOG measurements. However, a reanalysis of our present data using exactly the same parameters as in the previous study did not
change the present findings. Finally, MEG activity localized to the vmPFC is susceptible to several sources of variability that do not affect other cortical regions (such as the lATL and lAG) to nearly the same degree, and these difficulties might create greater instability in vmPFC effects. For example, it is has been demonstrated that eye movements can contaminate MEG measurements of neural activity from the vmPFC and surrounding regions (Carl et al. 2012). Though our results were not affected by the inclusion or exclusion of trials with excessive EOG amplitudes, subtle eye movements that fall beyond the ability of EOG to detect have been shown to appear as neural activity in both visual (Yuval-Greenberg et al. 2008) and auditory (Yuval-Greenberg and Deouell 2011) electrophysiological paradigms. It should be noted, however, that these distortions are not believed to be time-locked to the stimulus and have previously manifested themselves as induced rather than evoked effects. Also, as pointed out in our previous paper, the ability of MEG to reliably detect and localize cortical activity decreases with the distance from the sensors (Hillebrand and Barnes 2002). This limitation may contribute to more variable measurements from regions such as the vmPFC compared with locations closer to the surface, such as the lATL and lAG. So, in sum, there are several potential explanations for why the present results did not clearly replicate the combinatorial effect in the vmPFC that was previous observed, however, at the present time none of these theories unambiguously supply a complete explanation for this discrepancy. Clearly, more work is needed in order to disentangle how different neural generators contribute to later combinatorial effects.
On the other hand, there is now a growing body of converging evidence that implicates the lATL in basic linguistic combination. The present results further strengthen this conclusion and indicate that the cognitive operation subserved by this neural mechanism is utilized by both auditory and visual language processing.

*Modality-Specific Combinatorial Effects.*

For both modalities, we observed significant modality-specific combinatorial activity in cortical sources close to the primary cortices of the associated modality. During reading, we observed a large region of combinatorial activity located in the right occipital-temporal cortex. The location and timing (approximately 160ms following the onset of the noun) of this effect corresponds roughly to the secondary sensory response that is thought to occur following all visually complex stimuli – the M170 response (e.g. Tarkiainen et al. 1999; Solomyak and Marantz 2010). This component is hypothesized to reflect the analysis of orthographic and morphological properties of a visual word form, as it responds preferentially to strings of letters compared to symbols (Tarkiainen et al. 1999), correlates with properties of the visual form of a word (Solomyak and Marantz 2010), and reflects early morphological decomposition processes (Zweig and Pylkkänen 2009). Thus, our results suggest that in the present study basic combination in the visual modality was accompanied by increased processing related to the analysis of the visual form of the combined noun.

Similarly, during listening modality specific combinatorial sources were also identified close to the primary sensory cortex, this time centered over the left auditory cortex. Both the location, adjacent to the lAG ROI, and the timing of this effect,
concurrent with a marginally significant combinatorial effect in the lAG ROI at ~300ms, suggest that this effect reflects modality-specific combinatoric activity that occurs across a rather large cortical region encompassing both the primary auditory cortex as well as part or all of the lAG ROI, and precedes subsequent combinatorial activity in the latter region. A similar pattern of results was observed during a cross-modal comparison of anomaly processing (Hagoort and Brown 2000b) in which the N400 effect, observed in both modalities, was preceded by a topographically similar auditory-specific combinatorial effect. Taken together, these findings suggest that previous results reporting an earlier onset for auditory compared to visual combinatorial processing (e.g. Holcomb and Neville 1991; Osterhout and Holcomb 1993) might in fact reflect two separate auditory mechanisms – the first specific to auditory processing and the second shared between modalities (c.f. Hagoort and Brown 2000b). As previous MEG studies have generally localized processing of the acoustic form of an auditory word to a cortical region similar to the present result, (see Salmelin 2007), our results suggest that this first, auditory specific effect is associated with initial perceptual processing and that, as in visual processing, auditory combination in the current study was also accompanied by increased processing of the critical noun’s sensory form – in this case, its acoustic form.

Thus, at a general level, our modality specific findings were both relatively unsurprising and consistent across modalities. For both reading and listening, we found early combinatorial effects specific to each modality that were located near the associated primary and secondary sensory cortices and were reminiscent of past components thought to reflect the analysis of a word’s sensory form. The importance
of top-down, predictive information during the processing of combinatorial linguistic phrases has long been emphasized both theoretically (e.g. McClelland 1987; Hale 2003; Levy 2008) and empirically (e.g. Trueswell 1996; Spivey and Tanenhaus 1998; Gibson 2006). Additionally, top-down informational flow has been shown to modulate neural responses generated by the perceptual processing of both auditory (e.g. Debener et al. 2003) and visual (e.g. Engel et al. 2001) stimuli. Recently, MEG work investigating the interplay of these factors has shown that the early perceptual responses to a word can be modulated by the top-down prediction of its grammatical category (Dikker et al. 2009; Dikker et al. 2010). The results of the present study support this general framework and suggest that past differences observed between auditory and visual combinatorial effects might in part reflect top-down modulations of perceptual processing that appear to be active even during the comprehension of minimal linguistic phrases. Specifically, however, the functional significance of these modality-specific effects is unclear. It may be the case that the expectation of combination or the engagement of a grammatical parser causes more resources to be directed toward the initial processing of an impending word. However, at the present time, hypotheses regarding the specific mechanisms underlying these effects must remain conjecture until more targeted investigations directed specifically at this phenomenon have been carried out.

Conclusion.

Using a linguistically minimal paradigm, we were able to isolate combinatorial neural activity evoked by the construction of simple adjective-noun phrases in both speech
and reading. Further, we were able to do so using the same task and linguistic items across modalities. Our results indicate that linguistic processing in both modalities share common combinatorial mechanisms when constructing and understanding basic linguistic phrases. During both reading and listening, we found increased activity during basic composition in the IATL and the IAG. Further, in both cases the combinatorial IATL effect preceded combinatorial activity in the IAG. These results suggest that auditory and visual language processing draw upon the same neural mechanisms during basic combinatorial processing and that these mechanisms are deployed in a similar manner for both. Additionally, we identified modality-specific increases in combinatorial activity that temporally and spatially coincided with early perceptual responses, suggesting a modulation of early perceptual processing within both modalities during the comprehension of minimal linguistic phrases. At the present time, however, the exact functional significance of this modulation remains unclear. Further work must now build upon these results in order to determine how closely the correspondence between auditory and visual combinatorial neural mechanisms remains as the complexity of the stimulus increases.
References


Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME. 1988. Positron emission
tomographic studies of the cortical anatomy of single-word processing. Nature.
331(6157):585-589.

Phillips C, Pellathy T, Marantz A, Yellin E, Wexler K, Poeppel D, McGinnis M,

Poeppel D. 2003. The analysis of speech in different temporal integration windows:
cerebral lateralization as 'asymmetric sampling in time'. Speech
Communication. 41:245-255.

Poeppel D, Yellin E, Phillips C, Roberts TPL, Rowley HA, Wexler K, Marantz A.
1996. Task-induced asymmetry of the auditory evoked M100 neuromagnetic

Price CJ. 2000. The anatomy of language: contributions from functional

localisation of the visual and auditory word form areas: A reconsideration of

Pylkkänen L, Marantz A. 2003. Tracking the time course of word recognition with

Pylkkänen L, Stringfellow A, Marantz A. 2002. Neuromagnetic Evidence for the
Timing of Lexical Activation: An MEG Component Sensitive to Phonotactic
Probability but Not to Neighborhood Density. Brain and Language. 81:666-
678.


Figures

Figure 1. Experimental design. Our design crossed task (composition and list) and number of words (two and one). In each trial, participants indicated whether the target picture matched the preceding words. In the composition task, all preceding words were required to match the shape, whereas in the list task, any matching word sufficed. A total of six colors and 20 shapes were randomly combined and used as stimuli, and the trials were split evenly between matching and non-matching shapes. Only activity recorded at the final word (boat) was analyzed. All subjects performed both the auditory and visual version of the tasks, and the same linguistic materials were used for both sets of tasks.
Figure 2. Behavioral results. (A) Reaction time and (B) accuracy data were submitted to a 2x2x2 repeated-measures ANOVA with modality (visual vs. auditory), task (composition vs. list), and number of words (one vs. two) as factors. We observed a significant interaction between the task and number of words for reaction time (F(1,18) = 43.47), with post hoc tests revealing slower responses in the two-word list conditions and faster responses in the two-word composition conditions compared with matched one-word controls. For accuracy, there was a main effect of modality, with subjects more accurate in the auditory modality compared to the visual modality (F(1,18) = 8.753). Post-hoc within-task tests found that subjects were significantly more accurate on one-word trials compared to two word trials in every task except for visual composition, for which the two conditions did not differ significantly. Overall, the results indicate that two-word trials were harder than one-word trials for every task except visual composition, for which the opposite result held. *p<0.05; ***p <0.001.
Figure 3. MEG sensor data. The average evoked response to the critical noun is displayed for each modality, collapsed across conditions and subjects. (A) Following the presentation of a visual word, canonical evoked response peaks and field patterns (M100 and M170) (Tarkiainen et al. 1999; Pylkkänen and Marantz 2003) are visible at ~100 and ~150 ms, localized to the primary visual cortex and fusiform gyrus respectively. These components are followed by the M250 and M350 peaks and field patterns which localize to the left temporal lobe and are consistently observed during language comprehension (Embick et al. 2001; Pylkkänen et al. 2002). (B) Following the presentation of an auditory word, distinct M50 and M100 field patterns (Poeppel et al. 1996) can be seen, localized to the primary auditory cortex. These are also followed by the M250 and M350 components, again localized to the left temporal lobe.
Figure 4. Source analysis. (A) A full-brain map indicating the sources that were found to be preferentially active during combination for the different modalities. (B) The results of targeted ROI analyses on the activity localized to each region identified in the full-brain analysis. Rows represent activity from each ROI. Columns represent activity from each modality. Each pair of graphs represents activity from the Composition (green lines) and List (orange lines) tasks for a single ROI and modality.
Significant clusters of combinatorial activity are shaded in gray. In the Left ATL, we found significant combinatorial activity in both the visual (193-241ms; $p = 0.018$) and auditory (267-357ms; $p = 0.049$) modalities. In the Left AG, we found significant combinatorial activity in the auditory modality (537-591ms; $p = 0.040$) and marginally significant combinatorial activity in the visual modality (356-389ms; $p = 0.074$). (The light gray shading indicates a marginal combinatorial effect discussed further in the Results). Activity in a large portion of the right occipital-temporal cortex showed significant combinatorial activity in the visual modality (141-189ms; $p = 0.011$) and not the auditory modality, while activity near the primary auditory cortex displayed the opposite pattern of results (significant auditory cluster: 287-401ms; $p = 0.042$). No comparison within any List task showed any significant effect. ns, Nonsignificant; *p<0.05; **p <0.01.
**Chapter 3: Simple composition during language production in MEG** (Submitted to the *Journal of Neuroscience*)

**Abstract**

Historically, the neural basis of language production has been understudied relative to comprehension, due to difficulties inherent both in controlling the nature of the produced utterances and in measuring the associated brain responses. Nowhere is this disparity greater than in the study of combinatorial linguistic processing. While several recent studies have investigated the spatial dimension of neural processing associated with complex combinatorial productions such as sentences and narratives, the present study provides the first investigation into the temporal dynamics of basic combinatorial neural mechanisms that subserve language production. Using magnetoencephalography, we find that the production of simple adjective-noun phrases (‘red tree’) elicits increased neural activity that localizes to the ventromedial prefrontal cortex (vmPFC) and left anterior temporal lobe (LATL) relative to matched list controls (‘red, blue’), while the left inferior frontal gyrus (LIFG) is implicated in the reverse contrast. A previous MEG investigation (Bemis and Pylkkänen, 2011) also identified increased activity localized to both the LATL and vmPFC during the comprehension of these same minimal linguistic phrases, and further, the order of these effects was reversed in the two studies, with combinatorial vmPFC activity preceding LATL activity in the present study. Thus our results not only suggest that
the vmPFC and LATL play an integral role in basic composition during linguistic production, but also provide the first direct support that comprehension and production recruit a shared set of combinatorial neural mechanisms, operating in reverse order.
Introduction

Language provides the unique ability to transmit complex ideas from one person – the speaker – to another – the listener. Thus ultimately, the cognitive neuroscience of language must fully characterize the computations by which complex meanings are externalized as motor commands in language production and the way such meanings are constructed from sensory input in comprehension. Current research on the neurobiology of language, however, contributes to these goals highly asymmetrically as modern neuroimaging methods are used almost exclusively to study comprehension. This asymmetry is most pronounced for combinatory processing, as it is difficult both to elicit controlled combinatorial utterances from participants and to avoid contaminating motion artifacts as complex utterances unfold. The present work overcomes these challenges by combining the production of simple yet combinatory expressions with the millisecond time resolution of magnetoencephalography, providing a measure of compositional processing during a planning stage not yet contaminated by motion artifacts.

In hemodynamic methods, production artifacts are less severe than in electrophysiological paradigms, and so there exist a handful of imaging studies that address sentence production. This work has investigated language production globally, using narratives (Awad et al., 2007; Stephens et al., 2010), measured complexity effects, using words, phrases, and sentences (Indefrey et al., 2004), and identified neural circuits that underlie complex structural representations, using sentence adaptation (Menenti et al., 2011; Segaert et al., 2011). While this nascent body of
research is yielding important spatial information about the brain bases of combinatorial language production, characterizing the rapid temporal dynamics of linguistic computation will ultimately require a greater time resolution than fMRI or PET can afford. Unfortunately, techniques that match the speed of language (e.g., EEG and MEG) face great difficulty when investigating combinatory, multi-word productions as disentangling neural signals from muscle movements is extremely problematic once articulation begins. Thus, to date no electrophysiological paradigms have directly investigated the temporal dynamics of combinatorial neural mechanisms during language production.

Here we investigate the production of simple adjective-noun combinations (‘red tree’), whose semantic and syntactic combinatorial stages are thought to fully precede articulation (Alario et al., 2002; Schriefers et al., 1999). Thus, using MEG we can acquire a spatio-temporal map of these elusive processes with millisecond temporal accuracy and reasonable spatial resolution by contrasting their planning stages with those of matched, list controls (‘red, blue’). In a previous study (Bemis and Pylkkänen, 2011), we have shown that combinatorial processing during the comprehension of such phrases recruits the left anterior temporal lobe at ~250ms, followed by the ventromedial prefrontal cortex at ~400ms. Since production and comprehension are logical inverses, i.e., instead of semantic interpretation constituting the end product, constructing a “message” begins the computation (Bock and Levelt, 1994), it is reasonable to hypothesize that comprehension and production employ many of the same brain mechanisms and, at least to some extent, deploy them in reverse order (Indefrey and Levelt, 2004). Here we assessed whether cortical regions
previously associated with the comprehension of minimal phrases are also involved in their production, and whether their temporal order is reversed.
Materials and Methods

Participants.

Twenty non-colorblind, native English speakers participated in the study (12 female; mean age, 21.4 years). All provided informed written consent.

Task designs.

During the production task (Figure 1), participants either named colored shapes (‘red tree’) or two colors of a circular blob from left to right (‘red, blue’). Both tasks initially require the retrieval of two lexical items – hypothesized to occur largely in parallel for both elements, for both productions (Meyer et al., 2008; Schriefers et al., 1999) – followed by the phonological planning and articulation of these items (Ferreira and Engelhardt, 2007). Thus, we expect combinatorial mechanisms to evoke increased activity during phrase productions, and we expect to observe this activity prior to articulation (Alario et al., 2002).

In a visual control task (Figure 1), participants briefly viewed the same set of pictures and responded when they detected the presence of a linear luminosity gradient, which occurred on only ten percent of trials that were not included in further analysis. Since non-gradient trials provide a measure of neural activity evoked by simply viewing the production prompts without any associated linguistic processing or overt response, we expect to observe perceptual differences between the stimuli in both tasks, while linguistic effects should only manifest in the production task. To
minimize implicit naming, participants always performed the control task before the production task.

**Stimuli.**

Throughout the experiment, seven common one-syllable colors were used (black, blue, brown, green, pink, red, white) and each was assigned a corresponding object-denoting noun (cross, tree, star, glass, hand, bag, note) matched in phoneme number ($p > 0.35$), frequency ($p > 0.40$, HAL log frequency), and syllable structure. Two-color blobs were created from all 42 non-repeating combinations of colors, and colored shape trials were created by substituting each second color with its matched noun. Phrase prompts were hand created depictions of noun denotations, filled in with color, while list prompts were created by vertically dividing a circular blob into two colored halves. Linear luminosity gradients were applied using Photoshop, with the direction of the gradient aligned in one of four cardinal directions. To minimize eye movements, all pictures were presented foveally, subtending 1.33° onto a screen approximately 43cm from the participant’s eyes using psychtoolbox (Brainard, 1997).

During the production task, each stimulus was repeated three times, and trial types were blocked separately, with order randomized over five pairs of blocks. In the view task, each stimulus appeared twice and 16 additional stimuli contained gradients. All trial types were presented together in this task. Stimuli lists were randomized for each subject.

**Procedure.**
Before recording, we used a Polhemus Fastrak three-dimensional digitizer (Polhemus, VT, USA) to digitize participants’ head shapes, which were then used to constrain source localization during analysis by co-registering five coils located around the face with respect to the MEG sensors. During the experiment, participants lay in a dimly lit, magnetically shielded room. MEG data were collected using a whole-head 157-channel axial gradiometer system (Kanazawa Institute of Technology, Nonoichi, Japan) sampling at 1000Hz with a low-pass filter at 200Hz and a notch filter at 60Hz. To monitor blinks, the participant’s left-eye position was recorded at 1000Hz using an Eyelink-1000 (SR-Research, Osgoode, ON, Canada) infrared eye-tracking system.

Each trial initially presented a fixation for 400ms followed by the picture. In the view task, the picture remained for 300ms after which participants were given 1000ms to respond. In the production task, the picture remained until speech onset was detected using an individually adjusted volume threshold. Participants were then allotted 1200ms to respond. Vocal responses were recorded with an MEG compatible microphone (Optimic 1160, Optoacoustics, Or Yehuda, Israel) placed approximately 5cm from the participant’s mouth. Inter-trial intervals varied randomly, with an average [SD] of 200ms [50ms] in the view task and 500ms [100ms] in the production task. Participants completed the control task and then studied the intended color and object names prior to beginning the production task. The entire recording lasted ~40min.
Data acquisition.

Following past MEG picture naming paradigms (e.g. Levelt et al., 1998), to minimize movement artifacts, only activity during the preparation of productions, from 100ms before to 700ms after picture onset, was analyzed. In the view task, only non-gradient trials eliciting no response were analyzed, and one participant was eliminated for poor performance (ten standard deviations below average). Trials for which the maximum MEG amplitude exceeded 3000fT were removed along with those containing blinks, as identified by eye-tracking and supplemented with manual inspection of the MEG data. One further participant was removed due to excessively noisy data. Remaining data were then averaged for each participant for each condition and bandpass filtered between 1 and 40Hz.

Data analysis.

Minimum norm estimates of source activity were created for each average using BESA 5.1 (MEGIS Software GmbH). We assessed significant differences in localized activity using a targeted region of interest analysis, followed by a full brain comparison. For ROIs (Figure 2), we chose two regions previously implicated in basic combinatorial processing during the comprehension of minimal phrases – the left anterior temporal lobe (LATL) and ventromedial prefrontal cortex (vmPFC) (Bemis and Pylkkänen, 2011) – and two ROIs canonically associated with language production – the left inferior frontal gyrus (LIFG) and left posterior temporal lobe (LPTL) (Indefrey and Levelt, 2004). All ROI boundaries were unchanged from Bemis and Pylkkänen (2011). For each task, we assessed significant effects using a non-
parametric permutation test (Maris and Oostenveld, 2007) that identified temporal windows of significantly different activity between conditions, corrected for multiple comparisons across the entire epoch (see Bemis and Pylkkänen, 2011 for details). All reported $p$ values were obtained from distributions using 10,000 permutations of the original data.

Full-brain analyses compared activity values at every time-space data point within each task using a paired $t$-test. Differences were considered significant if they were reliable ($p < 0.05$) for at least 20ms and 15 adjacent cortical sources.
Results

Behavioral Results.

In the view task, there was no difference in accuracy between conditions (p > 0.35; average [SD]; colored shapes, 97.7% [1.8%]; two colors, 96.4% [5.8%]), though participants identified gradients significantly faster on two color blobs than colored shapes (p = 0.04; average [SD]; 259ms [112ms] vs. 311ms [87ms]), possibly due to a contrast between the linear gradient and the vertical separation between colors. The high overall accuracy, however, indicates that subjects paid attention to the stimuli during the task.

In the production task, phrase productions began significantly earlier than list productions (p = 0.002; average [SD]; 829ms [136ms] vs. 980ms [275ms]), possibly reflecting facilitated phrase processing, as observed during comprehension (c.f. Bemis and Pylkkänen, 2011), or inhibited list processing, perhaps due to semantic interference while retrieving the two color terms (c.f. Smith and Wheeldon, 2004). Importantly, however, response times in both conditions occurred sufficiently beyond the end of our analysis epoch to largely minimize contamination from muscle movements (Salmelin, 2007).
ROI results

Left anterior temporal lobe.

Two significant clusters of combinatorial activity were identified in the LATL (Figure 2) from 248-410ms ($p = 0.015$) and 423-620ms ($p = 0.009$). No increased activity was identified for list productions or for either condition in the view task (all clusters $p > 0.8$). The shortness of the gap between the two combinatorial clusters relative to their extents suggests a single, prolonged increase in LATL activity during the production of phrases compared to lists.

Ventromedial prefrontal cortex.

Two significant combinatorial clusters were also identified in the vmPFC (Figure 2) from 185-312ms ($p = 0.043$) and 520-630ms ($p = 0.039$). Again, no increased activity was observed during list productions or for either condition in the view task (all clusters $p > 0.8$). Thus, the cluster test identified combinatorial activity earlier in the vmPFC than the LATL (185ms v. 248ms), and a post hoc 2 x 2 repeated-measures ANOVA over averaged activity from 150-200ms revealed a marginally significant interaction between ROI and condition ($F(1,17) = 3.16$, $p = 0.09$).

Lateral inferior frontal gyrus and left posterior temporal lobe.

We observed no significant increases in LIFG or LPTL (Figure 2) activity during phrase productions or for either condition in the view task (all clusters $p > 0.8$). We did, however, observe a significant cluster of increased LIFG activity during list productions from 129-239ms ($p = 0.037$).
**Full-brain comparisons**

In general, our full-brain analyses (Figure 3) conform to our primary combinatorial ROI results. Clear combinatorial effects are visible in both the vmPFC (100-200ms) and the LATL (300-600ms). The lack of visible LIFG activity for list productions reflects the relatively weak and diffuse nature – in both space and time – of the underlying activity difference. Aggregating over both in the ROI analysis clarifies the effect. Outside of these regions, differences visible in the superior parietal lobule (phrases) and right inferior temporal cortex (lists) tentatively support previous results associating these regions with sentence generation (Haller et al., 2005) and color discrimination (Zeki and Marini, 1998), respectively.
Discussion

The present study represents the first investigation into the temporal dynamics of basic combinatorial brain mechanisms during language production. We used MEG to measure neural activity generated during the production of simple adjective-noun phrases as participants named colored shapes (‘red tree’). We then isolated combinatorial effects by contrasting this activity to that evoked as participants named two colors of a circular blob (‘red, blue’). Both productions are believed to initially involve similar, largely parallel retrievals of multiple lexical items followed by their phonological planning and articulation (Alario et al., 2002; Meyer et al., 2008). Thus, our design manipulates the presence of basic combinatorial production mechanisms while holding lexical and phonological demands largely constant. Further, because combinatorial processing of simple phrases is thought to complete prior to articulation (Alario et al., 2002), movement artifacts could be minimized by analyzing only activity generated between the presentation of the picture and the onset of speech. Finally, visual differences between the two prompt types were assessed through a non-linguistic, perceptual control task. Thus, our design isolates neural activity associated with basic combination during the production of minimal adjective-noun phrases and measures this activity using the millisecond temporal resolution of MEG. As previous electrophysiological production studies have focused exclusively on more complex utterances, the current study presents the first investigation into the temporal dynamics of basic combinatorial neural mechanisms that operate during language production.
Combinatorial LATL and vmPFC effects.

We observed significantly greater activity localized to both the LATL and vmPFC during the production of minimal phrases compared to lists, with significant combinatorial activity beginning earlier in the vmPFC than the LATL (~185ms vs. ~250ms after the presentation prompt). Increased activity persisted strongly in the LATL from ~300ms until 600ms, while combinatorial vmPFC activity was less consistent, though still present, throughout this interval as well. Neither region exhibited any significant effects during the perceptual control task. Thus, our results indicate that both the LATL and vmPFC support basic combinatorial processing during the production of minimal linguistic phrases and that combinatorial mechanisms subserved by the vmPFC precede those supported by the LATL.

Several hemodynamic studies have implicated both the LATL and the vmPFC in the production, and comprehension, of freeform autobiographical narratives using both spoken and signed language (Awad et al., 2007; Braun et al., 2001; Stephens et al., 2010). While narrative processing clearly relies upon a large number of diverse cognitive operations, the present results suggest that these effects reflect basic combinatorial mechanisms, which are clearly abundant during narrative production. Electrophysiologically, the few studies investigating multiword utterances have produced effects broadly consistent with ours, finding sustained differences beginning ~300ms following the production prompt (Habets et al., 2008; Marek et al., 2007), and, in one case, a frontally distributed differential from 180-230ms (Habets et al., 2008). Despite these temporal similarities, however, it is difficult to draw a direct correspondence from these findings to our own, as these studies explicitly contrasted
complex utterances with each other. Only one electrophysiological investigation has attempted to directly compare combinatorial and non-combinatorial productions (Eulitz et al., 2000), and they observed no significant differences between single word and noun phrase productions. However, in this study the critical comparison utilized covert production, which has been shown to elicit both muted and incomplete effects compared to overt production (Palmer et al., 2001), and thus few solid conclusions can be drawn from this null result. Thus, at a broad level, our results conform both to previous electrophysiological findings and to results derived from hemodynamic studies of narrative processing, however, neither set of studies offers a particularly close comparison to the present one as they investigate relatively complex utterances.

In general, few studies have investigated basic combinatorial linguistic processing in either production or comprehension. Our previous study, however, paradigmatically mirrored the current investigation and measured combinatorial neural activity evoked during the comprehension of simple adjective-noun phrases (Bemis and Pylkkänen, 2011). In this study, participants read such phrases (‘red tree’) and then judged whether a following colored shape matched the phrase. As in the present study, combinatorial effects were identified by contrasting this activity with that elicited by non-compositional list controls (‘cup, tree’), and, using identical analysis procedures, we also observed significantly increased activity in the LATL and vmPFC during basic combinatorial processing. Thus, combined, these two studies suggest that both the production and comprehension of minimal linguistic phrases recruit combinatorial mechanisms subserved by these two regions. Further, the temporal order of these effects was reversed in the two studies, with significant combinatorial
activity in the LATL preceding combinatorial vmPFC activity during comprehension. Though theoretical models rarely attempt to subsume comprehension and production under a single theory, those that do invariably posit shared semantic and syntactic mechanisms operating in reverse temporal order (e.g. Indefrey and Levelt, 2004). Empirically, however, evidence for shared combinatorial processing primarily derives from syntactic priming studies, which demonstrate that the comprehension of a specific type of syntactic expression affects subsequent production of that structure—an effect replicated both behaviorally (see Pickering and Ferreira, 2008 for a review) and neurally (Menenti et al., 2011). While these results suggest that syntactic representations are shared between modalities, their static nature prevents insight into whether specific combinatorial processes underlie both comprehension and production. Our study supplies the crucially important temporal dimension and suggests that the inverse relationship between production and comprehension posited in theoretical models is born out neurally through a shared set of basic combinatorial neural mechanisms. Further, the suggested delineation in Bemis and Pylkkänen (2011) of semantic and syntactic mechanisms to the vmPFC and LATL respectively is consistent with the present results and psycholinguistic models of picture naming, which posit lexical semantic mechanisms at ~175ms followed by syntactic processes at ~250ms (e.g. Indefrey and Levelt, 2004). Of course, further targeted investigations are needed before specific functional roles can be confidently assigned to either the vmPFC or LATL during basic linguistic combination.
Lack of combinatorial LPTL and LIFG effects.

We found no significant effects within the LPTL for any comparison. A wide range of hemodynamic and electrophysiological investigations implicate this region in the retrieval of phonological information during production (see Indefrey and Levelt, 2004 for a review). Thus, the lack of differential activity in this region suggests that phonological demands were held relatively constant between conditions, as desired.

In the LIFG, while we also observed no combinatorial effects, we did identify a significant cluster of increased activity during list productions, from 129-239ms. The LIFG, of course, has long been linked to language production, and recent neuroimaging studies associate activity in this region with the production of sentences (Dogil et al., 2002; Haller et al., 2005; Indefrey et al., 2004) and the representation of syntactic structure (Menenti et al., 2011; Segaert et al., 2011), supporting the hypothesis that this region encodes syntactic information during production (Indefrey et al., 2004). In the present study, however, we found no increased LIFG activity during the production of minimal syntactic phrases, but rather during the production of simple lists. This increase was accompanied by a delay in production onset, which suggests that it may be related to semantic interference arising from the retrieval of two semantically similar lexical items in this condition (‘red’, ‘blue’), as such interference is known to delay speech onset (Smith and Wheeldon, 2004) and has been associated with increased LIFG activity during production (Schnur et al., 2009). Additionally, previous electrophysiological evidence places semantic interference effects in the same time window as our LIFG effect (Maess et al., 2002), though localized more posteriorly. Thus, our results show no evidence that the LIFG
subserves basic combinatorial processing in the present manipulation and instead suggest that it is involved in selection among semantic competitors (c.f. Schnur et al., 2009).

**Conclusion.**

The present study fills a void in the current neurolinguistic literature by providing the first directed study into the temporal dynamics of basic combinatorial neural mechanisms that subserve language production. Rather than implicating traditional language regions, such as the LPTL ("Wernicke’s area") or the LIFG ("Broca’s area"), we find instead that basic combinatorial activity localizes to the LATL and vmPFC during the production of simple adjective-noun phrases (‘red tree’), compared to the production of matched list controls (‘red, blue’). A parallel study investigating the neural underpinnings of basic composition in comprehension (Bemis and Pylkkänen, 2011) identified increased activity in these same regions during the comprehension of the same minimal linguistic phrases. Thus, the present results not only suggest that these regions support basic combinatorial mechanisms in language production, but also that the supported mechanisms operate during both comprehension and production. Further, these effects were observed in the opposite temporal order during the two studies, supporting the widely held, but as of yet unsupported, contention that production and comprehension utilize the same neural pathway, but in reverse (Indefrey and Levelt, 2004). Future work is now needed in order to identify the precise functional structure of these generalized neural mechanisms that provide language with its combinatorial power.
References


Figures

Figure 1

A  **Production Task:** Say the phrase or list

B  **View Task:** Identify the gradients

**Figure 1.** In the production task (A), participants named colored shapes (‘red tree’) or two colors of a circular blob from left to right (‘red, blue’), with trial types blocked separately. In the view task (B), participants were given 1000ms to indicate if they detected a rare gradient on these same pictures, with trial types intermixed.
**Figure 2.** Localized activity is shown for the (A) LATL, (B) vmPFC, (C) LIFG, and (D) LPTL ROIs for both tasks, averaged across participants. Shaded regions indicate significantly greater activity during phrase production while the boxed region indicates significantly increased activity during list production, as determined using a non-parametric, permutation test (Maris and Oostenveld, 2007) performed over the entire epoch (10,000 permutations).
Figure 3. Full-brain analyses for the (A) view and (B) production tasks. Plotted regions denote amplitude differences for significant spatio-temporal effects ($p < 0.05$, uncorrected, for at least 20ms and 15 contiguous sources). Red denotes activity increases during phrase production (or colored shape viewing). Blue represents increases during list production (or colored blob viewing).
Chapter 4: Flexible Composition: MEG evidence for the deployment of basic combinatorial linguistic mechanisms in response to task demands

Abstract

Human language is defined by its combinatorial ability to flexibly construct an infinite number of complex meanings out of a finite number of individual pieces. The present study investigates whether a minimal manipulation in task demands can induce core combinatorial mechanisms that underlie this flexibility to extend beyond the bounds of normal grammatical phrases. Using MEG, we measured neural activity generated by the processing of simple adjective-noun phrases in canonical (*red cup*) and reversed word order (*cup red*). During a non-combinatorial task, subjects judged whether a following color blob and shape outline matched the linguistic items. During this task, we observed no significant combinatorial activity during the processing of reversed sequences, while canonical phrases elicited significant combinatorial activity – as assessed through activity localized to the LATL from 200-250ms (c.f. Bemis & Pylkkänen, 2011). When combinatorial demands were introduced into the task by simply combining the blob and outline into a single colored shape, significant combinatorial activity was observed during the processing of reversed sequences as well. These results demonstrate the first direct evidence that basic linguistic
combinatorial mechanisms can be deployed outside of their natural context in response to task demands, independent of lexical or attentional factors. This finding is discussed with respect to past investigations of linguistic task modulations as well the relationship between language and cognition more broadly.
Introduction

It is uncontroversial that human language derives its expressive power from the ability to flexible construct complex meanings out individual pieces. A recent theory, however, assigns an even more fundamental role to linguistic combinatorial mechanisms by suggesting that all uniquely human cognitive capabilities are derived from this process, and this process alone, as it allows for the combination of representations from otherwise informationally encapsulated, modular domains (Carruthers, 2002; Spelke, 2003). While the details of this theory remain to be fully worked out, the hypothesis raises an interesting question: Can basic combinatorial linguistic mechanisms – those that sit at the heart of language and compose complex meanings out of individual elements – be flexibly applied to novel contexts in response to task demands?

Clearly, there are many functional steps between the parsing of natural grammatical expressions and the combination of representations from diverse cognitive domains. In the present study, we seek to tackle the first of these by assessing whether task demands alone, independent of changes in lexical or attentional demands, can provoke the engagement of basic linguistic combinatorial mechanisms beyond their natural domain. To this end, in the present study, we manipulated task demands by only the amount necessary to modulate the relevance of combinatorial processing, and we investigated ‘novel’ contexts that differ from grammatical contexts by only the amount necessary to no longer naturally evoke combination between linguistic elements. Modulations in basic combinatorial linguistic processing were
then assessed during this manipulation by using magnetoencephalography to measure evoked neural activity within these contexts.

Few previous studies have directly investigated the extent to which combinatorial linguistic mechanisms can be flexibly deployed to novel contexts, though many results touch upon this question indirectly. Clearly, at some level, combinatorial linguistic processes can be applied to novel contexts, as people are capable of learning to extract meaning from written words and foreign languages, neither of which evoke successful combinatorial linguistic processing without instruction. Neurolinguistic investigations into the brain bases of both types of processing, while not definitive, have indicated a large overlap in the neural signatures associated with combinatorial speech comprehension and those evoked by both reading (Constable et al., 2004; Jobard, Vigneau, Mazoyer, & Tzourio-Mazoyer, 2007) and foreign language processing (Kotz, 2009 for a review; Perani et al., 1998; Rossi, Gugler, Friederici, & Hahne, 2006), though differences have also been observed for both (Dehaene et al., 1997; Michael, Keller, Carpenter, & Just, 2001), often correlated with proficiency (Perani et al., 1998; Rossi et al., 2006; Schlaggar et al., 2002). Moving beyond these investigations, a large body of evidence suggests that linguistic mechanisms can be extended to the processing of artificial ‘languages’ as well. In a canonical artificial grammar learning (AGL) paradigm (Reber, 1967), sets of arbitrary symbols, ranging from foreign words (Musso et al., 2003), to letter strings (Petersson, Forkstam, & Ingvar, 2004), to visual objects (Bahlmann, Schubotz, Mueller, Koester, & Friederici, 2009), are generated using a finite automaton such that they obey various syntactic constraints. Subjects are then shown exemplars from these
sets during a training period and later asked to judge the ‘grammaticality’ of a test set of strings, some of which are generated by the automaton and some of which are not. There are countless variations on the paradigm, in terms of syntactic constraints, ‘language’ symbols, learning method, and more (see Pothos, 2007 for a review), however, a consistent finding from both behavioral (Amato & MacDonald, 2010; Misyak, Christiansen, & Bruce Tomblin, 2010) and neural evidence (Friederici, Steinhauer, & Pfeifer, 2002; Musso et al., 2003; Petersson et al., 2004) is that learned artificial grammars are processed similarly to natural language largely to the extent that they embody syntactic constraints found in natural languages. Conversely, artificial languages constructed according to syntactic rules not found in natural language have been shown to no longer exhibit these similarities (Musso et al., 2003; Tettamanti et al., 2009). Thus, in broad strokes, AGL studies provide evidence that linguistic mechanisms can be deployed to novel contexts given appropriate task demands.

Nevertheless, several factors prevent results from AGL paradigms from establishing this fact conclusively. First, though subjects are usually able to distinguish ‘grammatical’ strings from ‘ungrammatical’ strings with better than chance accuracy (Pothos, 2007), their performance is rarely perfect (Fedor, Varga, & Szathmáry, 2012), and in some circumstances the underlying grammatical rules cannot be mastered at all, despite their apparent similarity to natural language constraints (M. H. de Vries, Monaghan, Knecht, & Zwitserlood, 2008; Perruchet & Rey, 2005). Further, because AGL paradigms are rooted in studies of language acquisition, the critical neural responses are often not observed until after an extensive learning period.
that can last for several days (Forkstam, Hagoort, Fernández, Ingvar, & Petersson, 2006; Friederici et al., 2002; Petersson, Folia, & Hagoort, 2010), though not always (Tettamanti et al., 2009). Thus, it is unclear at what speed and to what extent potential linguistic mechanisms are successfully deployed in these paradigms. Second, by design these studies almost exclusively measure processing associated with complex rules, such as hierarchal, nested structures (Bahlmann, Schubotz, & Friederici, 2008; Bahlmann et al., 2009) and non-rigid distances dependencies (Tettamanti et al., 2009), and are often focused on processing associated with the violation of these rules (e.g. Petersson et al., 2004). Thus, notorious difficulties in disentangling the myriad of mechanisms that underlie such processing in natural language (Grodzinsky & Santi, 2008; Willems & Hagoort, 2009) are exaggerated when dealing with artificial languages (c.f. Marcus, Vouloumanos, & Sag, 2003). Consequently, there has been extensive disagreement as to the nature of neural mechanisms that drive effects observed in AGL paradigms (M. de Vries, Christiansen, & Petersson, 2011; Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006; Friederici et al., 2002; Petersson et al., 2010; Petersson et al., 2004; Tettamanti et al., 2009) or even as to whether the putative rules have actually been learned at all (M. H. de Vries et al., 2008). Thus, while results from AGL paradigms suggest an ability to deploy linguistic mechanisms to novel contexts, their relationship to linguistic processes on the one hand and recursion (Fedor et al., 2012), expectation violation (Petersson et al., 2010), and hierarchical sequencing (Bahlmann et al., 2009; M. de Vries et al., 2011) mechanisms on the other has not been has not been entirely resolved.
Another difficulty in interpreting AGL results in the context of the current investigation is that the task ‘manipulation’ in these studies is both implicit and quite large – i.e. subjects are asked to draw upon implicitly learned complex constraints in order to intuit whether a string of symbols is ‘grammatical’ or not, and no comparison is usually made to activity generated within the same context but without this task (although see Friederici et al., 2002). As the present study endeavors to determine if a minimal task manipulation can engage combinatorial processing in novel contexts, we turn now to studies that explicitly investigate the effect of task demands on linguistic processing. By and large, however, most studies of this type have concentrated primarily on determining the extent to which linguistic processing can be limited when attention is altered in various ways and not on whether combinatorial mechanisms can be flexibly engaged in response to task demands. Such studies often direct attention away from the linguistic stimuli altogether, either passively, e.g. watching a silent movie while listening to speech (Hasting & Kotz, 2008; Shtyrov, Pulvermüller, Näätänen, & Ilmoniemi, 2003), or actively, e.g. performing an auditory discrimination task while listening to speech (Pulvermüller, Shtyrov, Hasting, & Carlyon, 2008), and those that do not divert attention away from the stimuli entirely are, for the most part, designed to direct attention to and from different aspects of the stimuli, e.g. by performing a font detection task (Gunter & Friederici, 1999) or selectively assessing syntactic or semantic violations (Rogalsky & Hickok, 2009). In general, these investigations have uncovered a somewhat intuitive gradient such that early processing stages appear to be largely invariant under attentional manipulations (Hahne & Friederici, 1999, 2002; Maidhof & Koelsch, 2011; Pulvermüller et al.,
mid-stage processing can often be modulated but not usually eliminated entirely (Deacon & Shelley-Tremblay, 2000 for a review; Holcomb, 1988; McCarthy & Nobre, 1993), and later processing can come and go depending on the task (Gunter & Friederici, 1999; Gunter, Friederici, & Schriefers, 2000; Hahne & Friederici, 1999; Hasting, 2008; Kuperberg, 2007 for a review).

This latter result potentially suggests the flexible deployment of linguistic mechanisms, and recent work has indeed indicated that later stages of linguistic parsing, often related to syntactic reanalysis and ambiguity resolution, are task dependent to some degree and do not always occur during normal language processing (Ferreira, 2003; Ferreira, Bailey, & Ferraro, 2002; Sanford & Sturt, 2002; Sturt, Sanford, Stewart, & Dawydiak, 2004). Such modulation, however, has been directly tied to the depth of processing of the stimuli – i.e. evidence indicates that comprehension need only be ‘good enough’ to solve the task – and a common measure of this processing, the P600 ERP component, seems to be influenced primarily by whether or not a judgment of plausibility is demanded by the given task (see Kuperberg, 2007). Thus, rather than reflect the flexible use of processes outside of their natural context, this evidence seems to indicate instead that parsing may halt before fully engaging mechanisms related to ambiguity resolution or syntactic reanalysis (Sanford & Sturt, 2002). To date, no evidence from these studies has directly addressed whether these mechanisms can be deployed outside of their natural context, excepting perhaps work on ungrammatical garden path sentences in which the parser can becomes confused (Phillips, Wagers, & Lau, 2011). Thus, rather than
suggesting the flexible engagement of linguistic processes, these results speak more to their automaticity.

In the present study, by contrast, we sought to investigate whether basic combinatorial linguistic mechanisms can be rapidly deployed outside of their natural contexts in response to task manipulations that explicitly do not accompany a change in complexity or attention. Specifically, we investigated whether the combinatorial mechanisms that serve to compose simple noun phrases such as red cup can also be recruited to compose the reversed sequence cup red, which does not conform to the English word order rules for adjectival modification.¹ In a non-combinatorial task, subject read these phrases alongside one-word controls and subsequently indicated whether a colored blob and a shape outline matched the linguistic stimuli (Figure 1). To engage combinatorial processing, we then simply combined the color and shape

¹ While noun-adjective sequences can be grammatical in English (e.g. wine red is a particular shade of red associated with wine), because our tasks require a judgment specifically regarding the shape denoted by the noun, we expect such construals to be minimized. Further, we selected combinations that were not familiar noun-adjective colors (e.g. beet red). Within larger contexts, adjectives may of course modify nouns post-nominally, e.g. I saw a cup red with paint, though such usage in general requires the adjective to be sufficiently ‘heavy,’ and often requires an intonational break following the noun (Sadler & Arnold, 1994). Further, it has been argued that even when grammatical, post-nominal modification differs semantically from prenominal modification in terms of the permanence of the property (Svenonius, 1993). The one class of exceptions to these generalizations – the modification of indefinite pronouns, e.g. something red – were avoided in the present study, as only specific object denoting nouns were used as stimuli.
into a single colored object, while maintaining the same linguistic stimuli and required judgment.

Our primary question was the extent to which reversed sequences (cup red) would engage combinatorial processing under these two task demands. To assess this, we presented each word sequentially and measured neural activity at the second word using magnetoencephalography (MEG). Previous results from this same paradigm have revealed reliable combinatorial activity during the processing of adjective-noun phrases localized to the left anterior temporal lobe (LATL) at $\sim$225ms and the ventromedial prefrontal cortex (vmPFC) at $\sim$400ms following the presentation of the second word (Bemis & Pylkkänen, 2011). In the present study, we utilize the earlier LATL effect as an index of basic combinatorial processing, as many electrophysiological studies suggest that later processing components, and specifically those elicited during the time window surrounding the vmPFC effect, can be heavily influenced by task manipulations, independent of any variation in combinatorial processing (see Deacon & Shelley-Tremblay, 2000). Additionally, the role of the LATL in combinatorial processing has been robustly supported by a vast hemodynamic literature (Friederici, Meyer, & Von Cramon, 2000; Humphries, Love, Swinney, & Hickok, 2005; Mazoyer et al., 1993; Vandenberghhe, Nobre, & Price, 2002), and so we expect to observe increased LATL activity during this earlier time window accompanying the deployment of combinatorial processing.

With respect to the canonical phrases, in principle, the non-combinatorial task could be carried out without any linguistic composition, via simply accessing colors and shapes. However, nearly all parsing models hypothesize that grammatical
linguistic expressions automatically engage combinatorial mechanisms (e.g. Frazier & Clifton Jr, 1989; MacDonald, Pearlmutter, & Seidenberg, 1994), and this intuitive claim has been supported by numerous neurolinguistic investigations demonstrating that early electrophysiological components reflecting combinatorial processing are invariant to both task (Hahne & Friederici, 1999, 2002) and attentional manipulations (Hasting, 2008; Pulvermüller et al., 2008; Shtyrov et al., 2003). Further, hemodynamic effects associated with syntactic manipulations remains observable even during tasks specifically designed to minimize linguistic processing (Caplan, 2010; Caplan, Chen, & Waters, 2008), and increased LATL activity remains robust during sentence processing even during explicitly non-combinatorial tasks (e.g. Vandenberghe et al., 2002). Consequently, we expected the canonical adjective-noun sequences to engage combinatoric routines even when not demanded by the task.

If our assumptions are correct and canonical adjective-noun combinations do engage combinatorial mechanisms in both tasks but reversed noun-adjective sequences do not automatically engage the same combinatorial mechanisms absent task demands, then we expect to see activity in the LATL exhibit an interaction in the non-combinatorial task between the reversed (cup red) and canonical orderings (red cup) compared to their matched one-word controls (xhl red and frw cup, respectively), driven by increased activity in the canonical order alone. Further, if basic combinatorial mechanisms are flexibly engaged by the combinatorial task during processing of the reverse sequences, we expect to see a main effect of number of words in this task, with increased LATL activity present in both two-word sequences relative to their matched controls. On the other hand, if this task manipulation is not
sufficient to evoke combinatorial processing, then we expect to observe the same interaction between order and number of words in this task as well. Note that because we measured activity on the second word, the canonical and reversed sequences cannot be compared directly, and the inclusion of the one-word comparisons is critical in controlling for lexical variation between these two conditions. Finally, to maximize the strength of the task manipulation, we administered each task separately to two different sets of subjects.
Methods and Materials

Participants.
15 subjects performed the non-combinatorial (Non-Compose) task (8 female; mean age = 22.4 years). 21 subjects performed the combinatorial (Compose) task (14 female; mean age = 21.4 years). All subjects were right-handed, non-colorblind native speakers of English with normal or corrected-to-normal vision. All gave written informed consent prior to the experiment and received a fee or course credit for their participation.

Materials.
Each trial contained a fixation cross, an initial word or non-word, a noun or adjective, and a target picture (Figure 1). Each stimulus was presented in isolation and followed by a blank screen. Subjects were told to ignore all non-word stimuli and indicate if the target picture contained a depiction of all of the preceding lexical items. In the Compose experiment, the target picture contained a single colored shape. In the Non-Compose experiment, the color and shape were presented separately as a circular blob and a white outline, respectively. Linguistic stimuli varied by condition: two-word canonical trials presented an adjective followed by a noun (red cup); two-word reverse trials presented a noun followed by an adjective (cup red); and one-word trials replaced initial words with unpronounceable consonant strings (xhl cup, frw red).
Thus, the second, critical stimulus remained unchanged between one and two-word trials.\(^2\)

Throughout all conditions, nine one-syllable, common color adjectives were used (red, tan, teal, blue, pink, black, brown, white, green). Each adjective was assigned a corresponding length-match noun (cup, car, lock, shoe, leaf, house, heart, plane, cross) and a corresponding length-match non-pronounceable consonant string (xkq, kjq, qxsw, mtpv, vbwnw, rjdnw, wvcnz, zbxlv, vitzkn). Nouns and adjectives were also matched for frequency ($p = 0.74$; HAL log frequency; paired $t$-test). Each word was displayed in Courier non-proportional font and subtended approximately 3°. In the Compose experiment, target pictures were hand-created canonical depictions of each shape, colored in with one of the nine colors and displayed in the center of the screen, subtending approximately 8°. In the Non-Compose experiment, circular blobs and outlines (taken from the colored shapes) were randomly placed at one of four locations, centered +/- 2° both horizontally and vertically from the center of the screen, with no two objects occupying the same location on any trial. In this experiment, each object subtended approximately 4° on its own. All stimuli were presented using psychtoolbox (Brainard, 1997; Pelli, 1997) and projected onto a screen ~45cm from the subject’s eye.

---

\(^2\) In the Non-Compose experiment, subjects also completed a list variant of each task, in which subjects were asked to judge whether the target picture matched either of two preceding nouns or adjectives. However, as this contrast is not relevant to the present discussion it has been omitted from the main body of the paper.
In each experiment, canonical and reversed trials were blocked separately, with order counter-balanced across each group of subjects. In each condition, critical items (i.e. nouns for canonical trials, adjective for reversed trials) were presented four times in matching trials and four times in non-matching trials, resulting in a total of 72 trials in each condition and 144 trials per block. In two-word non-matching trials, target pictures matched either the color or the shape, but not both. For each subject, two-word reverse trials were created by simply reversing the order of the stimuli in the two-word canonical trials. One-word trials were then created from all two-word trials by substituting a matched consonant string for each initial stimulus and shuffling the target pictures to match or not match appropriately. Stimuli lists were randomized per subject.

Procedure.

Before the experiment, subjects practiced their first task outside of the MEG room. Though subjects were made aware of the existence of further tasks at this time, no instructions or practice for subsequent tasks were given until after the completion of the first task. Prior to recording, subjects’ head shapes were digitized using a Polhemus Fastrak 3D digitizer (Polhemus, VT, USA). The digitized head shape was then used to constrain source localization during analysis by co-registering five coils located around the face with respect to the MEG sensors. Additionally, electrodes were attached 1 cm to the right of and 1 cm beneath the middle of the right eye in order to record the vertical and horizontal electrooculogram (EOG) and detect blinks. Both electrodes were referenced to the left mastoid.
MEG data were collected using a whole-head 157-channel axial gradiometer system (Kanazawa Institute of Technology, Tokyo, Japan) sampling at 1000Hz with a low-pass filter at 200Hz and a notch filter at 60Hz. All stimuli besides the target pictures were presented for 300ms, followed by a 300ms blank screen. Target pictures appeared at the end of each trial and remained onscreen until the subject made a decision. Subsequent trials began after a blank screen was shown for a variable amount of time following a normal distribution with a mean of 500ms and a standard deviation of 100ms (see Figure 1). The entire recording session lasted approximately 50 minutes in the Non-Compose experiment and 25 minutes in the Compose experiment.

**MEG Data Acquisition.**

MEG data from 100ms prior to the onset of each critical item to 600ms post onset were segmented out for each trial. Raw data were first cleaned of potential artifacts by rejecting trials for which the subject answered either incorrectly or too slowly (defined as more than 2.5 seconds after the appearance of the target shape), or for which the maximum amplitude exceeded 3000fT, or for which the subject blinked within the critical time window, as determined by manual inspection of the EOG recordings. One subject in the Non-Compose experiment performed at chance in the one-word reverse condition and so was excluded from further analysis. Remaining data were then averaged for each subject for each condition and band-pass filtered between 1 and 40Hz. For inclusion in further analysis, we required that subjects show a qualitatively canonical profile of evoked responses during the processing of the critical items. This
profile was defined as the appearance of robust and prominent initial sensory responses – i.e. either the M100 or M170 field pattern (e.g. Pylkkänen & Marantz, 2003; Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999) had to be clearly present in the time window of 100-200ms following the critical stimuli. Five subjects overall failed to meet this requirement (three in the Compose experiment and two in the Non-Compose experiment) and were excluded from further analysis.

Minimum Norm Estimates.

As our primary dependent measure, we created distributed minimum norm source estimates for recorded MEG sensor data. This measure provides an estimate of the cortical location of electrical activity underlying the observed magnetic fields recorded outside of the head. A source estimate was constructed for each condition average using L2 minimum norm estimates calculated in BESA 5.1 (MEGIS Software GmbH, Munich, Germany). The channel noise covariance matrix for each estimate was based upon the 100ms prior to the onset of the critical item in each condition average. Each minimum norm estimate was based on the activity of 1426 regional sources evenly distributed in two shells 10% and 30% below a smoothed standard brain surface. Regional sources in MEG can be regarded as sources with two single dipoles at the same location but with orthogonal orientations. The total activity of each regional source was then computed as the root mean square of the source activities of its two components. Pairs of dipoles at each location were first averaged and then the larger value from each source pair was chosen, creating 713 non-directional sources for which activation could be compared across subjects and conditions. Minimum norm
images were depth weighted as well as spatio-temporally weighted using a signal subspace correlation measure (Mosher & Leahy, 1998).

Data Analysis.

To assess the presence and modulation of combinatorial activity, our primary analysis examined source activity localized to a particular spatial-temporal region of interest, guided by results from our previous study (Bemis & Pylkkänen, 2011). Effects inside this spatial region but outside of the temporal window of this ROI were then examined within each task using a non-parametric permutation cluster test. Additionally, a full-brain analysis was performed within each task as well, comparing activity across all source and time points.

Our previous study was identical to the Compose experiment in the present study, except that two noun list sequences were used as a control task in place of the reversed sequences of the present experiment. For this condition subjects were asked to determine if the target picture matched either of the preceding words. In this study, we identified basic combinatorial activity associated with the comprehension of simple adjective-noun phrases that localized to the left anterior temporal lobe (LATL) from 184-255ms following the presentation of the noun. Thus, in the present study, we assessed combinatorial activity by averaging activity in each condition that was elicited from 200-250ms following the critical word that localized to the same LATL ROI as in Bemis & Pylkkänen (Bemis & Pylkkänen, 2011). Within each experiment, significant differences were then assessed using a 2 x 2 repeated-measures ANOVA with order (canonical v. reversed) and number of words (one v. two) as factors.
Following this targeted analysis, we searched for further differences in LATL activity between each set of two-word and one-word trials using a non-parametric cluster-based permutation test (Maris & Oostenveld, 2007) applied to the entire time interval between the onset of the critical item and the onset of the target picture. This test controls for multiple comparisons within the entire time interval and identifies clusters of time points for which activity is significantly different between conditions using the following procedure. First, a test statistic for the observed data is calculated by identifying contiguous time points for which the a paired $t$ test between conditions reaches a certain cutoff (set at $p = 0.30$ to match our previous analysis) and then summing the $t$ values of each of these tests at every time point in the cluster. Then, this same test statistic is calculated for many permutations of the actual data, created by randomly assigning the condition labels within each participant. Importantly, this assignment is done independently for each participant and only once per participant per permutation (as opposed to once for each time point for each participant). The $p$ value of the observed test statistic is then computed relative to a distribution created from 10,000 permutations of the original data and is set equal to the proportion of permuted datasets that produce a test statistic more extreme than that of the actual data.

To identify effects outside of our spatial ROI as well, we compared two-word and one-word activity measures within each task sample by sample for every source time point using a paired $t$ test. A difference was considered significant only if it surpassed a significance and size criteria such that it remained reliable ($p < 0.05$) for at least 10 samples (10 ms) and was observed in at least 10 adjacent cortical sources and
differed in amplitude by at least 1.5 nAm. In the results and figures below, we discuss only effects attributable to an increase in two-word activity compared with one-word activity for these comparisons.
Results

Behavioral results.

Accuracy and reaction time data for each experiment (Figure 2) were submitted to a 2x2 repeated-measures ANOVA with order (canonical v. reversed) and number of words (one v. two) as factors. Performance between the two experiments was then compared directly using a repeated-measures ANOVA employing a mixed-effects model with order (canonical v. reversed) and number of words (one v. two) as fixed effects and subject as a random effect.

Non-Compose behavioral results.

We observed no significant effects for reaction time for responses in the Non-Compose experiment for order, number of words, or their interaction (all $F$ values < 1; canonical: $p = 0.52$; two-words: 727ms avg. [86ms std.]; one-word: 739ms avg. [98ms std.]; reversed: $p = 0.96$; two-words: 759ms avg. [113ms std.]; one-word: 757ms avg. [116ms std.]). We did, however, observe a significant interaction between order and number of words for accuracy ($F(1,11) = 9.03, p = 0.012$). Follow-up paired $t$-tests reveal that this effect was driven by a significantly worse performance in the two-word canonical condition (red cup) compared to the matched one-word control ($p = 0.007$; two-words: 91.9% avg. [1.8% std.]; one-word: 96.8% avg. [0.7% std.]), and no statistical difference between accuracy in the reversed order conditions ($p = 0.39$; two-words: 97.9% avg. [0.8% std.]; one-word: 97.2% avg. [0.7% std.]).
**Compose experiment results.**

In the Compose experiment, we observed a significant interaction for reaction time between order and number of words ($F(1,17) = 8.39; p = 0.01$). Follow-up paired $t$-tests reveal that this effect was driven by significantly faster responses in the two-word canonical condition (*red cup*) compared to the matched control ($p < 0.001$; two-words: 579ms avg. [30ms std.]; one-word: 644ms avg. [38ms std.]) and no statistical difference between the reversed order conditions ($p = 0.40$; two-words: 625ms avg. [36ms std.]; one-word: 640ms avg. [33ms std.]). For accuracy, we observed no interaction between number of words and order and no main effect of order (both $F$ values $< 1$). We did, however, observe a significant main effect of number of words ($F(1,17) = 13.94; p = 0.0017$). Follow-up paired $t$-tests revealed significantly lower accuracy for two-word canonical responses compared to their paired one-word controls ($p = 0.003$; two-words: 96.4% avg. [0.8% std.]; one-word: 98.1% avg. [0.5% std.]) while accuracy in the reversed condition only showed a trending difference ($p = 0.015$; two-words: 97.1% avg. [0.4% std.]; one-word: 98.0% avg. [0.4% std.]).

**Comparison between experiments.**

For accuracy, we observed a significant three-way interaction between task (Compose v. Non-Compose), order (canonical v. reversed), and number of words (two v. one) ($F(1,28) = 9.95; p = 0.004$) confirming that the change in task significantly affected the interaction between number of words and order in this measure. Specifically, this result indicates that the two-word reversed condition in the Non-Compose task patterned similarly the four one-word results and not like the other two-word results.
For reaction time, the three-way interaction was not significant \( F(1,28) = 1.44, p = 0.28 \) though the interaction between number of words and task was marginally significant \( F(1,28) = 3.56, p = 0.07 \), driven by faster reaction times for two-word responses in the Compose task compared to the Non-Compose task. Though not as clear an effect as for accuracy, this result does support the conclusion that the two-word reversed condition in the Non-Compose experiment patterned separately from the other two-word conditions as there was also a significant interaction between number of words and order \( F(1,28) = 4.75; p = 0.038 \), driven by faster responses overall for the two-word canonical conditions. Thus, the marginal interaction between task and number of words indicates faster response times for the two-word conditions in the Compose task and the significant interaction between number of words and order indicates faster response times for the two-word canonical conditions in both tasks. Thus, only the two-word reversed responses in the Non-Compose experiment were not involved in any differences relative to the one-word controls. However, the lack of significance for the three-way interaction means that this evidence is merely suggestive with respect to response times.

Nevertheless, there appears to be strong behavioral evidence that differences between one-word and two-word responses patterned similarly across all comparisons except the reversed order in the Non-Compose task. In this comparison, no differences were observed between one-word and two-word responses. While the cause of the decrease in accuracy for two-word responses in the other three comparisons is not clear, the faster response times exhibited in all three of these conditions echoes the facilitation previously observed for composed phrases in this task (Bemis &
Pylkkänen, 2011) and others (Graves, Binder, Desai, Conant, & Seidenberg, 2010; Potter & Faulconer, 1979). Thus, in sum we believe that our behavioral results indicate the processing of reversed sequences in the Compose task patterned with canonical adjective-noun processing, while the processing of reversed sequences in the Non-Compose task patterned with one-word responses.

Left anterior temporal lobe ROI results.
In each experiment, activity localized to the LATL (Figure 3) was averaged from 200-250ms for each condition and submitted to a 2x2 repeated-measures ANOVA with order (canonical v. reversed) and number of words (one v. two) as factors. Activity between the two experiments was then compared directly using a repeated-measures ANOVA employing a mixed-effects model with order (canonical v. reversed) and number of words (one v. two) as fixed effects and subject as a random effect. Additionally, significant differences in LATL activity outside of this time window were investigated within each task using a non-parametric cluster test applied to the entire time interval between the presentation of the critical word and the target picture.

Non-Compose LATL results.
We observed a significant interaction between order and number of words for activity localized to the LATL from 200-250ms in the Non-Compose experiment ($F(1,11) = 5.72; p = 0.0358$). As expected, this effect was driven by significantly greater activity in the two-word canonical condition compared to the one-word control ($p = 0.01$; two-words: 4.15nAm avg. [0.61nAm std.]; one-word: 2.89nAm avg. [0.40nAm std.]) and
no difference between the two reversed conditions ($p = 0.80$; two-words: 3.78nAm avg. [0.51nAm std.]; one-word: 3.63nAm avg. [0.39nAm std.]). Subsequent cluster tests over the entire interval from 0-600ms revealed only a cluster of significantly greater two-word compared to one-word activity in the canonical conditions from 194-296ms ($p = 0.0357$), thus confirming the results of the targeted analysis.

*Compose LATL results.*

In the Compose experiment, we observed no interaction between order and number of words for activity localized to the LATL from 200-250ms ($F < 1$). We also observed no main effect of order ($F < 1$). We did, however, observe a significant main effect of number of words ($F(1,17) = 4.51, p = 0.0486$). While increased activity did not reach significance within either of the paired two-word and one-word comparisons (reversed: $p = 0.13$; two-words: 4.32nAm avg. [0.60nAm std.]; one-word: 3.52nAm avg. [0.44nAm std.]; canonical: $p = 0.21$; two-words: 3.86nAm avg. [0.40nAm std.]; one-word: 3.30nAm avg. [0.25nAm std.]), a targeted cluster test over a slightly larger time window, from 200-300ms, identified a cluster of significantly greater two-word activity compared to one-word activity in the reverse conditions from 214-279ms ($p = 0.0481$). A corresponding cluster test within the canonical conditions did not find any significant clusters of increased activity though, due apparently to a slight dip in increased two-word activity in the middle of this interval which resulted in two separate clusters of increased activity from 201-216ms and 245-263ms, neither of which were significant on their own ($p = 0.27$ for both). Subsequent cluster tests applied to the entire time interval did not reveal any further significant effects for
either order, though a targeted post-hoc analysis of activity at the visually salient peak in the two-word canonical condition did show increased activity relative to the one-word control from 475-525ms ($p = 0.03$; paired $t$ test). A similar test over the peak in the two-word reverse condition from 125 to 175ms showed no significant effect ($p = 0.11$).

Comparison between experiments.
The three-way interaction between task, order, and number of words failed to reach significance for localized LATL activity from 200-250ms ($F(1,28) = 2.195; p = 0.15$), however, a slightly tighter interval in the middle of this epoch, from 220-240ms, did show a significant three-way interaction ($F(1,28) = 4.337; p = 0.047$), and a wider interval from 200-300ms demonstrated a marginally significant three-way interaction ($F(1,28) = 3.853; p = 0.060$).

Thus, put together our results strongly indicate that combinatorial LATL activity from 200-250ms occurred in both canonical two-word conditions and in the two-word reverse sequences during the Compose task. We found no evidence of any increased activity at all in this ROI for the two-word reverse condition in the Non-Compose experiment. Thus, overall, our results exhibit the pattern predicted to arise from the flexible deployment of basic combinatorial mechanisms to reversed sequences during the Compose task.

Full-brain results.
In general, both full-brain analyses (Figure 4) support our ROI analysis quite closely, with clearly visible increases in LATL activity during two-word processing for both canonical orderings and for reverse sequences in the Compose task. Further, no LATL activity is visible during the processing of reverse two-word sequences in the Non-Compose experiment. A more detailed discussion is given below for each experiment.

*Non-Compose full-brain results.*

For the canonical ordering, a clear increase in activity can be seen at 150-250ms in the LATL, conforming to the ROI analysis above. Additional effects are also visible in the ventromedial prefrontal cortex (vmPFC) and right anterior temporal lobe (RATL) from 450-550ms. Increased activity in both of these regions was observed during the processing of adjective-noun phrases in our previous study, though both began slightly earlier and only vmPFC activity appeared to truly reflect combinatorial processing (see Bemis & Pylkkänen, 2011 for details). A follow-up post-hoc analysis within the vmPFC ROI from the previous study however failed to show any significant effects, as did a more circumscribed ROI constructed from the present full-brain analysis. Thus, results from this region echo many earlier studies that demonstrate task modulations for electrophysiological components in this time window (Deacon & Shelley-Tremblay, 2000).

Within the reversed sequences, we observed very few increases in activity during two-word processing, which supports our previous ROI and behavioral results in suggesting that processing in the two-word reverse condition patterned similarly to the one-word condition in this contrast.
Compose full-brain results. For the canonical ordering, by far the largest increase in activity during two-word processing occurred in the LATL and extended from 250-550ms. This result supports the above ROI analysis in indicating that significant increased activity occurred in the LATL during two-word canonical processing in this task. Further, it suggests that the relative weakness of the effect observed within the predetermined spatio-temporal ROI follows from the unexpectedly extended nature of this effect. As previously mentioned, this condition does not differ in any structural way from that in our previous study which elicited increased LATL activity from ~200-250ms followed by increased vmPFC activity from ~300-500ms; a result replicated in the present study for the Non-Compose task. Thus, the different pattern of results in the present contrast, with later increased activity occurring in the LATL as opposed to the vmPFC, is somewhat unexpected. One possible explanation might be that the present study used only nine different critical nouns whereas the previous study employed 25. This decrease in lexical variability might have contributed to the muted nature of the later effects in the present contrast, as decreased lexical variability has often been associated with decreased neural activity (e.g. Gabrieli et al., 1996). However, more directed work is clearly needed before this conclusion can be endorsed in any serious way. Outside of activity in the LATL, the only other visible effect in this contrast occurs in the RATL, which, as discussed above, was also observed in our previous study.

In the reversed conditions, the full-brain results again conform to our ROI analysis and show a clear increase in LATL activity from 150-250ms. Effects visible
outside of this ROI closely align with our previous findings for canonical adjective-noun phrases (Bemis & Pylkkänen, 2011) and indicate increased RATL activity early in the epoch followed by a later increase in vmPFC activity. A post-hoc ROI analysis of this latter activity reveals a robust cluster of increased activity very similar to that observed in our previous study, both in terms of temporal placement (377-478ms in the present study; 331-48ms in the previous study) and strength (both $p < 0.01$). Thus, the similarity between these results and those observed during adjective-noun processing in our previous study further underscores the conclusion that combinatorial processing extended to the processing of reversed sequences in this task.
Discussion

In the present study, we investigated whether basic combinatorial linguistic processing can be flexibly deployed to a novel context given only a minimal change in task demands. We recorded MEG activity as subjects read simple adjective noun sequences in canonical or reversed order (*red cup* v. *cup red*), and we identified combinatorial processing by assessing increases in activity localized to the LATL following the presentation of the second word compared to that evoked by matched, one-word controls (*xhl cup, frw red*). When subjects were required to judge whether the words matched a following colored blob and shape outline, no combinatorial processing was observed for reversed sequences (*cup red*). When the following target was instead presented as a single colored shape, significant combinatorial processing was observed for these same sequences. Further, behavioral measures indicated consistent differences between two-word and one-word processing, with the exception of two-word reversed sequences in the Non-Compose experiment, which patterned with the one-word conditions. Thus, the present data, both behavioral and neural, indicates that basic combinatorial linguistic mechanisms can be flexibly engaged outside of their normal context given only a minimal change in task demands.

Task modulations and linguistic processing.

Past investigations into the effects of task demands on language processing have focused nearly exclusively on manipulations of attention, either explicitly (Pulvermüller et al., 2008) or implicitly (Hasting & Kotz, 2008). Results from these studies indicate that linguistic processing exhibits a gradated response profile such that
early linguistic processes appear highly robust to many types of attentional diversions (Hahne & Friederici, 1999, 2002; Pulvermüller et al., 2008) while later processing becomes progressively more malleable (Deacon & Shelley-Tremblay, 2000; Kuperberg, 2007). None of these studies, however, directly investigate whether this later malleability belies an ability to flexibly deploy linguistic mechanisms outside of their normal contexts, instead remaining focused on how changes in attention affect the degree to which combinatorial mechanisms are engaged during the course of normal linguistic processing.

Even the few studies that do employ tasks that appear targeted at modulating combinatorial processing remain explicitly focused on attentional effects. For example, a previous investigation (Graves et al., 2010) employed an fMRI design very similar to our own in which subjects read familiar noun-noun compounds in both canonical and reversed order (lake house v. house lake). Subjects then either performed a ‘non-combinatorial’ task – identifying if a word was repeated from one presentation to the next – or a ‘combinatorial’ task – determining if a given sequence was ‘meaningful’ or not. Despite the well-known complexities involved in processing noun-noun combinations (c.f. Murphy, 2002), this paradigm might have served to investigate the deployment of combinatorial mechanisms at large had the ‘combinatorial’ task not been explicitly structured to disregard all evidence of such processing. While it remains somewhat unclear exactly how reversed noun-noun sequences (house lake) might or might not be meaningful – as opposed to less familiar or more ambiguous in meaning – in this study the authors chose to define all reversed sequences as ‘non-meaningful’ a priori, regardless of the subject’s response. Thus, all
trials that might have revealed the deployment of combinatorial mechanisms, i.e. those for which subjects reported constructing a meaning for the reversed sequences during the combinatorial task, were disregarded in the analysis, and the authors explicitly stated that they were interested only in assessing increases in combinatorial processing driven by increased attention to meaning in the ‘meaningful’ compounds.

In direct contrast, the present study sought to minimize changes in attention while modulating only combinatorial processing. In both tasks, subjects had to retrieve the meaning of all words in order to determine whether the target picture contained their denotation and thus, processing related to lexical access and attention should be equated between the two tasks. Therefore, it is difficult to attribute the increase in LATL activity observed during the processing of reversed sequences in the Compose experiment to changes in attention, especially as the same activity was robustly present for canonical phrases during the Non-Compose experiment as well. Thus, unlike past investigations, the present study demonstrates the flexible engagement of combinatorial linguistic mechanisms evoked by a change in task demands that does not simultaneously modulate attention. While this result is consistent with the hypothesis that previously observed processing modulations, such as those that underlie theories of ‘good enough’ processing, reflect the flexible deployment of combinatorial mechanisms, the relationship between such observations and the present finding is not entirely clear.

Different types of combinatorial processing.
Likewise, it is not completely clear how the present results bear upon those arising from AGL paradigms. While both behavioral (Amato & MacDonald, 2010; Misyak et al., 2010) and neural evidence (Friederici et al., 2002; Musso et al., 2003) suggest that artificial grammars can be processed similarly to natural language, these results are usually obtained after relatively extensive training with the grammar (Forkstam et al., 2006; Petersson et al., 2010) that, even then, only produces performance that is ‘significantly greater than chance’ and far from that of natural language (Fedor et al., 2012). Further, these studies focus almost exclusively upon the processing of complex syntactic structures, thus making it extremely difficult to determine whether their results indicate the involvement of linguistic mechanisms (Friederici et al., 2002; Musso et al., 2003) or more general processes that handle hierarchical sequencing more generally (Bahlmann et al., 2009; M. de Vries et al., 2011). Contrastingly, in the present study we observed extremely high accuracy overall and manipulated processing within simple two-word sequences. Thus, though our results are certainly consistent with the hypothesis that artificial grammars can engage linguistic mechanisms, differences in speed, success, and complexity between the present study and AGL paradigms prevents a particularly solid connection from being drawn at the present time.

It should be noted though, that while the present study does not employ complex linguistic expressions, this in and of itself does not enable an easy determination of the specific linguistic mechanisms that were engaged; it merely narrows the scope of possibilities. Even the most basic combinatorial linguistic processing evokes both syntactic and semantic mechanisms, responsible for forming
structural relationships and constructing complex meanings from individual elements. Previous work has most prominently linked hemodynamic activity in the LATL to syntactic mechanisms during combinatorial linguistic processing, as activity in this region has been shown to decrease during syntactic priming (Noppeney & Price, 2004) and increase with structural complexity (Brennan et al., 2012), however, other findings suggest that activity in this region might reflect semantic combinatorial processing as well (Pallier, Devauchelle, & Dehaene, 2011; Rogalsky & Hickok, 2009).

While the nature of our task manipulation might appear to indicate that semantic combinatorial demands were modulated – by requiring the formation of a singular mental conception of a colored shape to judge against the Compose target – past results suggest that semantic combination might only proceed after syntactic processes have completed (e.g. Hah02). Thus, though semantic combination may have been the impetus for combinatorial processing in the present study, syntactic combination might have been obligatorily engaged as well. In potential support of this hypothesis, many past results associate early, automatic electrophysiological components with syntactic processes (Hahne & Friederici, 2002; Hasting, 2008; Pulvermüller et al., 2008) – but see (Hinojosa, Martín-Loeches, Muñoz, Casado, & Pozo, 2004) – which suggests a syntactic role for the LATL in the current study, as early, robust activity in this region was observed during the processing of canonical phrases in the Non-Compose task despite the explicitly non-combinatorial nature of the task. Further, in our past work, we observed significant combinatorial activity in the vmPFC following increased LATL activity (Bemis & Pylkkänen, 2011), while in the current study activity in this region is much muted during similar processing in the
Non-Compose task. This activity profile – a late occurring processing component that can be modulated by task demands – matches that of many other electrophysiological effects primarily associated with semantic processing (Deacon & Shelley-Tremblay, 2000; Holcomb, 1988), and past MEG evidence has linked activity in this region with semantic composition (Pylkkänen & McElree, 2007). Thus, the present results are consistent with a syntactic role for the LATL and a semantic role for the vmPFC, however, since no attempt was made to explicitly disentangle these two intertwined processes, this interpretation is hardly ironclad.

*Relation to investigations of domain-generality.*

Though the present study is clearly not directly about non-linguistic processing, the results seem germane to the discussion as many past investigations into the relationship between linguistic and non-linguistic processing can be construed as suggesting a flexible deployment of combinatorial linguistic processing to a novel context. ERP components and hemodynamic responses canonically observed during language processing have also been evoked by a wide range of putatively non-linguistic stimuli, such as pictures (Nigam, Hoffman, & Simons, 1992; West & Holcomb, 2002; Willems, Özyürek, & Hagoort, 2008), actions (Özyürek, Willems, Kita, & Hagoort, 2007), music (Besson & Macar, 1987; Koelsch, Gunter, Wittfoth, & Sammler, 2005; Patel, Gibson, Ratner, Bessel, & Holcomb, 1998), mathematical sequences (Niedeggen, Rösler, & Jost, 1999), gestures (Kelly, Ward, Creigh, & Bartolotti, 2007), environmental sounds (Van Petten & Rheinfelder, 1995) and movies (Reid & Striano, 2008; Sitnikova, Kuperberg, & Holcomb, 2003). Interestingly,
however, none of these results have been interpreted as demonstrating the extension of linguistic mechanisms to other cognitive domains. Rather these findings have uniformly been offered as evidence that domain-general mechanisms underlie linguistic processing. In other words, shared effects between language and other cognitive domains are invariably interpreted as supporting the existence of a domain-general mechanism that underlies both types of processing, as opposed to demonstrating that linguistic mechanisms can operate in a different domain. In fact, even when a purely linguistic interpretation is maintained for a seemingly shared effect, instead of suggesting that the linguistic mechanism has been extended into the other cognitive domain, an appeal is usually made to functional heterogeneity or non-overlapping but similar cortical regions (e.g. Grodzinsky & Santi, 2008).

On the one hand, the findings put forward here can be viewed as a ‘proof of concept’ that basic linguistic mechanisms might be flexibly deployed into contexts that do not initially evoke such processing. On the other hand, it is also quite possible that a domain-general mechanism underlies basic linguistic combinatorial operations and that syntactic cues inherent in canonical phrases can override task demands that would otherwise dissuade such processing. Distinguishing between these two hypotheses in any experiment is extremely difficult, even in theory, and most likely will depend upon either developmental or evolutionary contrasts with non-linguistic subjects (Hauser, Chomsky, & Fitch, 2002), though even then the evidence may not be easy to sort out (Marcus, 2006). Nevertheless, the present results hopefully serve to raise the profile of this alternative hypothesis and, in light of current theories that place linguistic combinatorial abilities at the center of all human cognition (Spelke, 2003),
suggest that future investigations into the relationship between language and cognition focus not only on how cognition serves to enable language, but also on how language serves to enable cognition.

Conclusion.

In the present study, we demonstrate that basic combinatorial linguistic mechanisms can be flexibly deployed to a context beyond their natural domain. In contrast to previous studies that manipulate attention or utilize lengthy implicit learning paradigms, we show here that merely introducing the relevance of combination into a task, while holding attentional and lexical factors constant, is sufficient to elicit basic combinatorial processing between two lexical items that do not naturally engage such processing. Future work can now build on this result in order to determine the extent of this flexibility both in terms of linguistic mechanisms and cognitive domains.
References


203


Figures

**Figure 1: Experimental Design**

![Experimental Design Diagram]

**Figure 1**: Experimental design. In each block of trials, subjects were presented with both one-word and two-word stimuli and asked to judge if a following target picture matched the preceding words. In the canonical conditions (A), stimuli were adjective-noun phrases (*red cup*) and their matched one-word controls (*xhl cup*). In the reversed conditions (B), stimuli were noun-adjective sequences (*cup red*) and their matched one-word controls (*frw red*). In the Non-Compose task (C), target pictures contained a colored blob and a shape outline. In the Compose task (D), target pictures contained a single colored shape. All subjects completed one block of each sequence order in only one task.
Figure 2: Behavioral Results

A) Non-Compose

**Accuracy**

<table>
<thead>
<tr>
<th></th>
<th>Canonical</th>
<th>Reversed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td><img src="image" alt="Accuracy" /></td>
<td><img src="image" alt="Accuracy" /></td>
</tr>
</tbody>
</table>

**RT**

<table>
<thead>
<tr>
<th></th>
<th>Canonical</th>
<th>Reversed</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td><img src="image" alt="RT" /></td>
<td><img src="image" alt="RT" /></td>
</tr>
</tbody>
</table>

B) Compose

**Accuracy**

<table>
<thead>
<tr>
<th></th>
<th>Canonical</th>
<th>Reversed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td><img src="image" alt="Accuracy" /></td>
<td><img src="image" alt="Accuracy" /></td>
</tr>
</tbody>
</table>

**RT**

<table>
<thead>
<tr>
<th></th>
<th>Canonical</th>
<th>Reversed</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td><img src="image" alt="RT" /></td>
<td><img src="image" alt="RT" /></td>
</tr>
</tbody>
</table>

Two-word Canonical  Two-word Reversed  One-word Canonical  One-word Reversed

Figure 2: Behavioral results. For each task, reaction time and accuracy data were submitted to a 2 x 2 repeated-measures ANOVA with order (canonical vs. reversed) and number of words (one vs. two) as factors. In the Non-Compose task (A), we observed a significant interaction between the two factors for accuracy (F(1,11) = 9.03, p = 0.012), with post hoc tests revealing lower accuracy for the two-word canonical condition compared with the matched one-word controls. No significant
effects were found for reaction time in this task. In the Compose task (B), we observed a significant main effect of number of words for accuracy (F(1,17) = 13.94; p = 0.0017), with *post hoc* tests revealing significantly lower accuracy for two-word canonical responses compared to their matched one-word controls. We also observed a significant interaction between the two factors for reaction time (F(1,17) = 8.39; p = 0.01), with *post hoc* tests revealing significantly faster responses for the two-word canonical condition compared with the matched one-word controls. ns, Nonsignificant; ***p < 0.001; **p < 0.01.
Figure 3: LATL ROI Results. Localized activity is shown for the LATL ROI during the processing of the critical items (nouns in canonical sequences, adjectives in reversed sequences), averaged across subjects. In the Non-Compose task (A), we observed a significant interaction between order and number of words for activity measured from 200-250ms \((F(1,11) = 5.72; p = 0.0358)\). In the Compose task (B), we
observed a significant main effect of number of words (F(1,17) = 4.51, p = 0.0486) for activity measured from 200-250ms.

**Figure 4: Full-brain Results**

A) Non-Compose

<table>
<thead>
<tr>
<th>Canonical (red cup – xhl cup)</th>
<th>Reversed (cup red – frw red)</th>
</tr>
</thead>
<tbody>
<tr>
<td>msec following cup</td>
<td></td>
</tr>
</tbody>
</table>

B) Compose

<table>
<thead>
<tr>
<th>Canonical (red cup – xhl cup)</th>
<th>Reversed (cup red – frw red)</th>
</tr>
</thead>
<tbody>
<tr>
<td>msec following cup</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4: Full-brain Results.** Plotted regions denote the difference in average amplitude between two-word and one-word conditions for all space time regions in which two-word activity was reliably greater than one-word activity (p < 0.05, uncorrected) for at least 10 ms over 10 spatial neighbors, and the amplitude differed by at least 1.5 nAm between conditions. For clarity, non-cortical sources have been removed. Results within the Non-Compose task (A) conform to our ROI analysis and
reveal a clear LATL effect from 150-250ms for the canonical sequences. A slight RATL and vmPFC effect are also observable at 450-500ms. No clear effects are visible in the reversed comparison. Results within the Compose task (B) also conform to our ROI analysis and show a clear LATL effect in both the canonical and reversed sequences. Additional increased activity can also be seen in the vmPFC and RATL at 150-250ms and 350-450ms, respectively.