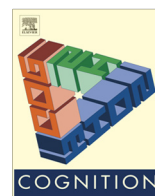




ELSEVIER

Contents lists available at ScienceDirect

## Cognition

journal homepage: [www.elsevier.com/locate/COGNIT](http://www.elsevier.com/locate/COGNIT)

# Building phrases in language production: An MEG study of simple composition



Liina Pyllkkänen<sup>a,b,c,\*</sup>, Douglas K. Bemis<sup>a</sup>, Estibaliz Blanco Elorrieta<sup>c</sup>

<sup>a</sup> Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA

<sup>b</sup> Department of Linguistics, New York University, 10 Washington Place, New York, NY 10003, USA

<sup>c</sup> NYUAD Institute, New York University Abu Dhabi, PO Box 129188, Abu Dhabi, UAE

## ARTICLE INFO

### Article history:

Received 26 November 2012

Revised 21 May 2014

Accepted 7 July 2014

Available online 14 August 2014

### Keywords:

Language production

Magnetoencephalography

Anterior temporal lobe

Ventro-medial prefrontal cortex

Left inferior frontal gyrus

Minimal phrases

## ABSTRACT

Although research on language production has developed detailed maps of the brain basis of single word production in both time and space, little is known about the spatiotemporal dynamics of the processes that combine individual words into larger representations during production. Studying composition in production is challenging due to difficulties both in controlling produced utterances and in measuring the associated brain responses. Here, we circumvent both problems using a minimal composition paradigm combined with the high temporal resolution of magnetoencephalography (MEG). With MEG, we measured the planning stages of simple adjective–noun phrases ('red tree'), matched list controls ('red, blue'), and individual nouns ('tree') and adjectives ('red'), with results indicating combinatorial processing in the ventro-medial prefrontal cortex (vmPFC) and left anterior temporal lobe (LATL), two regions previously implicated for the comprehension of similar phrases. These effects began relatively quickly (~180 ms) after the presentation of a production prompt, suggesting that combination commences with initial lexical access. Further, while in comprehension, vmPFC effects have followed LATL effects, in this production paradigm vmPFC effects occurred mostly in parallel with LATL effects, suggesting that a late process in comprehension is an early process in production. Thus, our results provide a novel neural bridge between psycholinguistic models of comprehension and production that posit functionally similar combinatorial mechanisms operating in reversed order.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

Language expresses thought by externalizing linguistic expressions built from internal conceptual representations. Currently, however, little is known about the neural implementation and temporal arrangement of core linguistic combinatorial processes that construct complex utterances from simpler pieces. Difficulties inherent in both eliciting controlled multi-word productions and avoiding

contaminating motion artifacts have limited neural investigations into combinatorial production mechanisms and have rendered electrophysiological studies, along with the temporal information they convey, virtually non-existent. The present work provides an initial investigation into this critical missing dimension, overcoming these obstacles by combining simple yet combinatory expressions with the millisecond time resolution of magnetoencephalography to yield a measure of compositional processing not yet contaminated by motion artifacts.

Research employing hemodynamic methods, being less susceptible to motion artifacts than electrophysiological paradigms, has begun to explore the neural bases of

\* Corresponding author. Address: New York University, 10 Washington Place, New York, NY 10003, USA. Tel.: +1 (212) 998 8764.

E-mail address: [liina.pyllkkanen@nyu.edu](mailto:liina.pyllkkanen@nyu.edu) (L. Pyllkkänen).

combinatorial language production (e.g. Awad, Warren, Scott, Turkheimer, & Wise, 2007; Indefrey, Hellwig, Herzog, Seitz, & Hagoort, 2004; Menenti, Gierhan, Segaert, & Hagoort, 2011), however these results lack fine-grained temporal information about observed neural activity. Unfortunately, faster electrophysiological techniques are rather problematic for the investigation of complex productions, as disentangling neural signals from muscle movements is extremely difficult once articulation begins. Consequently, electrophysiological paradigms have concentrated almost exclusively on single word utterances (Indefrey & Levelt, 2004; Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998; Salmelin, 2007) with the few exceptions investigating conceptualization difficulty as opposed to linguistic combinatorial operations (Habets, Jansma, & Münte, 2008; Marek, Habets, Jansma, Nager, & Münte, 2007). Only one electrophysiological investigation has attempted to isolate more fine-grained combinatorial processes (Eulitz, Hauk, & Cohen, 2000), however no significant effects associated with combinatorial productions were observed. This manipulation, however, relied on covert production, which can elicit both muted and incomplete effects compared to overt production (Palmer et al., 2001). Thus, to date, the temporal dynamics of basic combinatorial neural mechanisms that underlie language production remain largely uncharted.

### 1.1. Our study

In the present study, we directly investigate these processes by using MEG to measure neural activity generated during the production of simple adjective–noun combinations ('red tree') as subjects named pictures of colored shapes. Critically, behavioral evidence indicates that conceptual and grammatical encoding for such productions is completed before articulation begins (Alario, Costa, & Caramazza, 2002; Meyer, 1996; Schriefers, De Ruiter, & Steigerwald, 1999). Therefore, by measuring activity following the production prompt but preceding speech onset, we can obtain a spatio-temporal map of combinatorial mechanisms, such as the encoding of conceptual semantic relationships and structural syntactic relationships between elements (Ferreira & Slevc, 2007; Levelt, 1989), uncontaminated by movement artifacts.

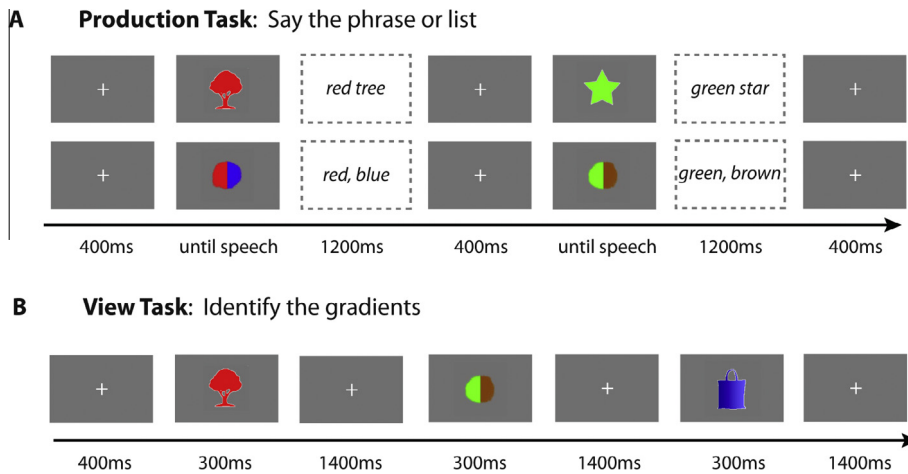
Perhaps the most challenging design question in a production study aiming to address basic composition is that of an appropriate non-combinatory control condition. Ideally, the control condition should engage similar mental processes as the composition condition, excepting only the critical combinatory operations. It should elicit similar vocal onset times as the combinatory stimulus, or at least not faster ones, as then the motion artifacts of the control condition might be concurrent with the critical combinatory activities of the phrase condition, prohibiting a comparison. Thus one-word productions are problematic controls at least in an initial study, as they are produced faster than two word productions. Further, since adjective–noun combinations of the sort described above ('red tree') are highly natural productions, the control productions should also be reasonably natural, i.e., expressions that people might naturally utter in life outside of the

lab. Otherwise the control condition might be excessively difficult, eliciting globally higher amplitudes than the more natural phrase and thus swamping any combinatory effects in the composition condition. This consideration rules out the otherwise attractive option<sup>1</sup> of asking subjects to generate the critical phrases in reverse order as the control condition, i.e., 'tree red.' While this control would match all lexical and phonological factors with the phrases, we took the cost of having subjects generate ungrammatical utterances to be too high for this matching. Further, in our work on basic composition during comprehension, we have found that whether or not reversed sequences such as 'tree red' elicit combinatorial effects similar to their canonical word order variants ('red tree') is highly task dependent, making this type of choice even riskier for an initial production investigation (Bemis & Pyllkkänen, 2013).

With these considerations in mind, we chose to compare adjective–noun phrase productions to productions of two color names generated in a list-like fashion: 'red, blue'. These expressions were elicited by asking subjects to name the colors of a circular blob from left to right (Fig. 1). Lists are of course not obviously non-combinatorial; in the context of the present task, the lists were semantically equivalent to the conjunction of the two colors (cf. e.g., Zimmermann, 2000). However, the list productions lacked the critical computation of combining the two words into a single concept, and thus our study should be taken as more narrowly focusing on this specific process. In our work on basic composition in comprehension, we have shown that the comprehension of adjective–noun phrases such as 'red boat' is associated with a specific profile of activation not elicited for disjunctively interpreted lists ('cup, boat' within a task eliciting a 'cup or boat' type interpretation) (Bemis & Pyllkkänen, 2011), providing prior evidence that lists are processed qualitatively differently from adjectival modification of nouns. Critically, the lists in the current study matched the phrases on several lexical factors that strongly influence processing during prearticulation – number of produced words (Schriefers et al., 1999), phonological onset (Cholin, Levelt, & Schiller, 2006), and word frequency (Oldfield & Wingfield, 1965). Finally, behavioral evidence indicates that in simple lists, as in phrases, lexical retrieval of both items precedes articulation (Malpass & Meyer, 2010; Meyer, Ouellet, & Häcker, 2008).

In sum, our goal was to investigate the profile of activation elicited by productions involving the adjectival modification of nouns, as compared to list-like productions of two colors. Effects due to visual differences between conditions were assessed with a separate control task (Fig. 1). Additionally, after having characterized a profile of activation associated with adjective–noun phrases vs. adjective–adjective lists, we conducted an additional control experiment examining the extent to which these effects may be attributable to syntactic category as opposed to composition, i.e., to the presence of a noun in the phrasal condition only (Experiment 2).

<sup>1</sup> Suggested by reviewers.



**Fig. 1.** Experiment 1 design. In the production task (A), subjects named either colored shapes ('red tree') or two colors of a circular blob from left to right ('red, blue'). In the control view task (B), subjects indicated if they detected a rare gradient on these same pictures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

## 1.2. Predictions

Though past electrophysiological evidence does not supply a clear prediction for when combinatorial effects should occur during phrase production, psycholinguistic models suggest that syntactic information becomes accessible with the lemma (Ferreira, 2000; Levelt, 1989). Thus, we expect combinatorial operations to initiate during the retrieval of conceptual and lexical information, placed at ~175–250 ms by single-word electrophysiological studies (Indefrey & Levelt, 2004). Spatially, recent neuroimaging studies supply a substantial array of cortical regions potentially associated with combinatorial language production, such as the left anterior temporal lobe (LATL), left inferior frontal gyrus (LIFG), left superior temporal gyrus (LSTG), and anterior cingulate cortex (ACC) (Awad et al., 2007; Braun, Guillemin, Hoesy, & Varga, 2001; Dogil et al., 2002; Haller, Radue, Erb, Grodd, & Kircher, 2005; Indefrey et al., 2004). However, the differences between these studies and the present paradigm in both neural measure and linguistic complexity preclude straightforward predictions. Only one paradigm has investigated similarly simple expressions, identifying increased LIFG activity during the production of adjective–noun phrases compared to word lists in PET (Indefrey et al., 2001). If this finding reflects a basic role for the LIFG in combinatorial production, then we would expect to see a similar effect in the present contrast as well.

The strongest predictions however follow from the proposal that comprehension and production share combinatorial neural mechanisms (e.g. Indefrey & Levelt, 2004). As already mentioned, our prior work contrasted MEG activity generated during the *comprehension* of minimal adjective–noun phrases ('red tree') with non-compositional list controls (Bemis & Pykkänen, 2011), constituting a close parallel to the current production design. Our results revealed two primary effects associated with the adjectival modification of nouns: a LATL increase in an early time window (~200–250 ms) followed by increased

activity in the ventro-medial prefrontal cortex (vmPFC) at ~350–450 ms. This LATL effect straightforwardly conforms to a large hemodynamic literature reporting increases in left anterior temporal activity for structured sentences as compared to unstructured lists (Friederici, Meyer, & von Cramon, 2000; Humphries, Binder, Medler, & Liebenthal, 2006; Humphries, Love, Swinney, & Hickok, 2005; Mazoyer et al., 1993; Rogalsky & Hickok, 2009; Snijders et al., 2009; Stowe et al., 1998; Xu, Kemeny, Park, Frattali, & Braun, 2005), which in a sense, is a sentence-level version of the minimal composition paradigm of Bemis and Pykkänen (2011). The ventromedially localizing prefrontal effect was somewhat more surprising in light of prior hemodynamic literature, but had a strong precedent in our own previous MEG work, where this source location has systematically shown effects of semantic complexity both for sentences (Brennan & Pykkänen, 2008, 2010; Pykkänen & McElree, 2007; Pykkänen, Martin, McElree, & Smart, 2009; Pykkänen, Oliveri, & Smart, 2009; for a review, see Pykkänen, Brennan, & Bemis, 2011) and pictures (Bemis & Pykkänen, 2012a), suggesting a semantic mechanism that extends beyond language (cf., Roy, Shohamy, & Wager, 2012). In contrast, combinatorial effects in the LATL have not extended to pictures (Bemis & Pykkänen, 2012a), suggesting a possibly more language specific computation.

The relative paucity of ventromedially localizing effects in the hemodynamic language literature could result from many factors (for a detailed review, see Pykkänen et al., 2011). For example, MEG is likely to be more sensitive than fMRI to orbitofrontally localizing activity, given that fMRI susceptibility artifacts make medial prefrontal regions notoriously difficult to image (Ojemann et al., 1997). Consistent with this, though the vmPFC has not figured in hemodynamic sentence vs. word list studies, an MEG version of this design implicated both the LATL and the vmPFC as loci of increased activity for sentences (Brennan & Pykkänen, 2012). A slightly different methodological consideration arises from the fact that the vmPFC is also a

strong contributor to the brain's default network, i.e., a set of regions active during free thought or "rest" (Buckner, Andrews-Hanna, & Schacter, 2008; Gusnard & Raichle, 2001). Thus whenever a control task allows for more "mind wandering" than the test task, it is possible that vmPFC effects during the test task can be cancelled out. This could be the case for the sentence vs. word list contrast, as unstructured word lists are rather boring stimuli do not require much attention. In fact, one special property of the Brennan and Pykkänen (2012) sentence vs. word list study in MEG was that the experimental task on the word lists was a somewhat difficult recall task, which may have suppressed default mode activity during list reading and allowed for vmPFC effects to be detectable for the sentences. Overall, because semantic processing is intuitively a component of the brain's resting state, it is not surprising that the default network and the "semantic network" are largely overlapping, both crucially containing the vmPFC (Binder & Desai, 2011; Binder, Desai, Graves, & Conant, 2009). Other hemodynamic evidence conforming to the vmPFC combinatory effects in MEG include increases in ventromedial orbitofrontal activity in PET when the comprehension of a story increases (Maguire, Frith, & Morris, 1999) or when subjects are asked to complete sentences with a word that fits the context as opposed to completing a sentence with a word that does not fit (Nathaniel-James & Frith, 2002). Further, the vmPFC participates in a network of areas showing speaker–listener coupling in fMRI during successful communication (Stephens, Silbert, & Hasson, 2010) and emerges as a shared region for narrative comprehension and production in a large scale review spanning imaging and neuropsychological studies (Mar, 2004). In sum, the combination of the evidence that narrative level vmPFC effects cover both comprehension and production and the evidence that in comprehension, vmPFC composition effects are elicited even for minimal two-word phrases (Bemis & Pykkänen, 2011) sets up a strong prediction that the production of small phrases should also engage the vmPFC.

To conclude, the primary aim of our study was to test whether the LATL and vmPFC effects observed for the comprehension of adjective–noun phrases extend to production, a prediction arising from any model where production and comprehension use the same combinatory machinery. Given that production is at least in some sense the reverse of comprehension, of particular interest was the temporal ordering of potential LATL and vmPFC effects, which in comprehension were timed LATL first (200–250 ms), then vmPFC (~400 ms) (Bemis & Pykkänen, 2011). For completeness, the LIFG was also included as a region of interest in our analyses, given the prior results of Indefrey et al. (2001).

## 2. Experiment 1

### 2.1. Materials and methods

#### 2.1.1. Participants

Twenty non-colorblind, right-handed, native English speakers participated (12 female; 21.4 years average). All

were neurologically intact, with normal or corrected-to-normal vision and all provided informed written consent.

#### 2.1.2. Stimuli

The production task (Fig. 1), consisted of 126 trials on which the subjects named a colored shape ('red tree') and 126 trials on which they named the two colors of a circular blob from left to right ('red, blue'). Seven adjectives (*black, blue, brown, green, pink, red, white*) and seven nouns (*cross, tree, star, glass, hand, bag, note*) were used. To achieve a well-controlled stimulus set, the nouns and adjectives were created pair-wise and matched on phoneme number (color words: 3.71 mean, [0.49 sd]; shape words: 3.57 mean, [0.53 sd];  $p = 0.35$ , paired  $t$ -test), and log HAL frequency (color words: 11.2 mean, [0.88 sd]; shape words: 10.7 mean, [8.7 sd];  $p = 0.43$ , paired  $t$ -test) (Lund & Burgess, 1996). Two-color blobs were created for all 42 non-repeating color combinations, and adjective–noun combinations were constructed by substituting the second color of each pair with its matched noun. Picture prompts were hand-created depictions of each utterance. Since the task instruction was slightly different for the two types of productions (phrases vs. lists), they were presented in separate blocks with block order randomized over five block pairs.

Since the seven colors were used in both first and second position, the list productions involved more repetition than the phrasal productions. Since composition is typically a facilitatory factor for response time (Bemis & Pykkänen, 2011; Potter & Faulconer, 1979), this repetition was likely to draw the phrase and list productions closer together in terms of response times, a desirable design feature for measuring the planning stages prior to production. However, despite any facilitation from repetition, we replicated the typically observed speed-up for the phrasal stimuli compared to the list controls (see Results). Thus, the list stimuli were clearly harder, biasing the experiment against observing increased activity during the phrasal productions, at least under the (rather standard but not obviously true) assumption that harder processing recruits more brain activity.

In the view task, subjects viewed the same pictures as in the production task and pressed a button if they detected the presence of a rare luminosity gradient. Trials with no gradient required no response. Linear luminosity gradients were applied using Photoshop, with the direction of the gradient aligned in one of four cardinal directions. All pictures were presented foveally using psychtoolbox (Brainard, 1997) and subtended  $1.33^\circ$  on a screen ~43 cm from the subject. During the production task, each stimulus was repeated three times. In the view task, each stimulus appeared twice, resulting in 86 colored-shape trials and 86 circular blob trials. An additional 16 randomly selected pictures contained gradients. Since the task was uniform for both trial types in the view task, stimulus order was randomized (no blocking) and subjects simply paused to rest twice during the task. The view task was always completed before the production task.

#### 2.1.3. Procedure

Before recording, each subject's head shape was digitized using a Polhemus Fastrak three-dimensional digitizer



(Polhemus, VT, USA). MEG data were collected using a using a whole-head 157-channel axial gradiometer system (Kanazawa Institute of Technology, Kanazawa, Japan) as subjects lay in a dimly lit, magnetically shielded room. Blinks were monitored using an Eyelink-1000 (SR-Research, Osgoode, ON, Canada) infrared eye-tracking system. Vocal responses were recorded with an MEG compatible microphone (Optimic 1160, Optoacoustics, Or Yehuda, Israel).

In both tasks, trials began with a fixation cross (400 ms), followed by a picture prompt. In the production task, the picture remained onscreen until speech onset began. Subjects were then given 1200 ms to finish speaking. The timing of the view task was designed to yield a profile of activation for early visual processing (before 300 ms) while adding as little time to the overall recording session as possible. For this reason, the picture onscreen for only 300 ms in this task. Subjects were then given 1000 ms to respond. After picture offset, subjects were likely to blink, and thus data in this task were only analyzed for the 0–300 ms interval. Inter-trial intervals varied according to a normal distribution (view: 200 ms mean [50 ms sd]; production: 500 ms mean [100 ms sd]). The entire recording lasted ~40 min.

#### 2.1.4. Data acquisition

Eye-tracking data were recorded at 1000 Hz and monitored the position of the subject's left eye. Due to calibration problems, usable eye-tracking data was however only obtained for 13 participants (though one of these was later removed for excessive noise in the MEG data). MEG data were recorded at 1000 Hz (200 Hz low-pass filter, 60 Hz notch filter) and epoched from 100 ms before to 700 ms after picture onset. In the view task, 2.8% of the trials [3.3% sd] were removed due to incorrect behavioral responses (one subject with accuracy 10 standard deviations below the mean was eliminated). We also removed trials with MEG amplitudes exceeding 3000 fT, and those containing blinks, as identified by eye-tracking data when available and supplemental manual inspection of the MEG data. Altogether, this resulted in the exclusion of 51.0% of the trials (14.1% sd) in the production task, leaving 123.4 trials on average per subject (35.5 sd), and 39.6% (9.9% sd) in the view task, leaving 101.5 trials on average per subject (16.6 sd). Thus a relatively high proportion of trials were lost due to movement artifacts and blinks, however our design included a sufficient number of trials to remain robust to this loss and result in a standard numbers of trials for inclusion in the analysis. One subject, however, did not display this robustness and was removed due to excessive noise.

#### 2.1.5. Statistical analysis

Preprocessed data were averaged for each condition and subject and bandpass filtered (1–40 Hz). For each subject condition average, the 100 ms prior to the onset of the critical item served as the baseline period for the epoch. L2 Minimum norm estimates of source activity were then created for each average using BESA 5.1 (MEGIS Software GmbH). Our primary analysis investigated activity localized to specific regions of interest: the LATL, vmPFC, and

LIFG. As our goal was to examine to what extent the combinatory LATL and vmPFC effects of Bemis and Pykkänen (2011) replicate in production, our statistical analysis closely followed this prior study. For each task and region, a non-parametric permutation test (Maris & Oostenveld, 2007) was used to identify temporal clusters during which activity localized to each region differed significantly between conditions, corrected for multiple comparisons. Importantly, this test controls for multiple comparisons across the entire analysis window and can therefore be used without precise *a priori* temporal hypotheses. As explained in Maris and Oostenveld (2007), in this test, criteria are first set for selecting clusters to be evaluated in the permutation test, i.e., waveform separations that last for a given amount of time at a given alpha-level (here, 10 adjacent time points with  $p < 0.3$ , as in Bemis & Pykkänen, 2011, where these thresholds were chosen to identify relatively shallow but long-lasting patterns of activity). Then, a test statistic is constructed for each cluster. In this case, as in Bemis & Pykkänen, 2011, this statistic was equal to the summed *t*-values of the point-by-point test-statistics over the selected cluster interval. Finally, the observed data are permuted by randomly assigning condition labels within each participant's data a given number of times (here 10,000 as in Bemis & Pykkänen, 2011), and the final corrected *p*-value of the observed data is calculated as the ratio of permutations yielding a test statistic greater than the actual observed test statistic. To determine the significance of this final value, the standard alpha-level of  $p < 0.05$  was used. As in Bemis and Pykkänen (2011), the LATL ROI covered BA 38 and the anterior portions of BA 20 and 21, the vmPFC ROI BA 11 bilaterally and the LIFG BAs 44 and 45 (precise source coordinates available on request).

The ROI analysis was followed by an uncorrected full-brain analysis to verify that the effects observed in the ROI analysis in fact reflected activity centered within the ROIs and not spill-off from neighboring regions. This analysis also enabled us to detect possible effects localized outside of our ROIs. In this full-brain analysis, activity values were compared at every time-space data point using a paired samples *t*-test. The problem of multiple comparisons was alleviated by only considering effects significant if they remained reliable ( $p < 0.05$ ) for at least 20 ms and 15 adjacent cortical sources.

## 2.2. Results

### 2.2.1. Behavioral results

In the view task, subjects identified gradients significantly faster for blobs than shapes (259 ms mean [112 ms sd] v. 311 ms mean [87 ms sd],  $t(17) = 2.25$ ,  $p = 0.038$ , two-tailed), possibly due to a contrast between the linear gradient and the vertical separation between colors. Accuracy did not differ (shape: 97.7% mean, [1.8% sd]; blobs: 96.4% mean [5.8% sd]),  $t(17) = 0.92$ ,  $p = 0.37$ , two-tailed). Phrase productions began significantly earlier than lists (829 ms mean [136 ms sd] v. 980 ms mean [275 ms sd],  $t(17) = 3.49$ ,  $p = 0.003$ , two-tailed), possibly reflecting facilitated phrase processing, as observed during comprehension (Bemis & Pykkänen, 2011; Potter &

Faulconer, 1979), or inhibited list processing, potentially due to semantic interference between the color terms (Smith & Wheeldon, 2004). During the recording sessions, it was clear that accuracy in the production task was at ceiling, with each participant making approximately one error in the course of the whole experiment. For this reason we forewent analyses of accuracy and rejections of incorrect trials from the MEG data in this task.

### 2.2.2. ROI results

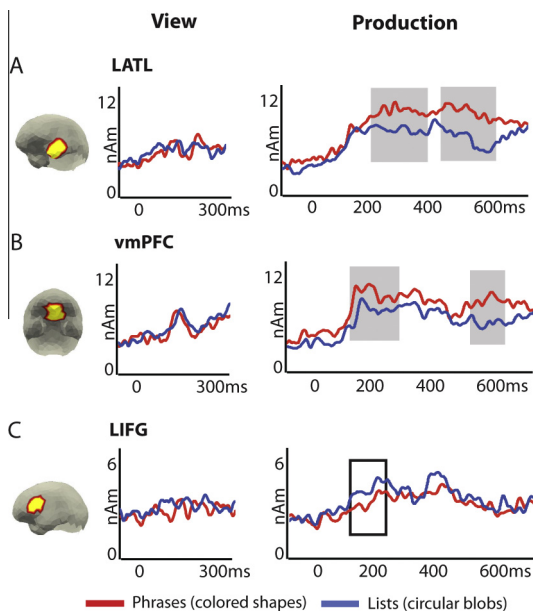
In the production task, we observed (Fig. 2) two significant clusters in the LATL in which localized activity was significantly greater during Phrases compared to Lists (248–410 ms;  $p = 0.015$ ; 10,000 permutations; 423–620 ms;  $p = 0.009$ ; 10,000 permutations). During both, localized activity was significantly greater during phrases compared to lists (cluster 1: 9.39 nAm Phrase mean [4.88 nAm sd] v. 7.20 nAm List mean, [4.82 nAm sd]; cluster 2: 9.20 nAm Phrase mean [5.99 nAm sd] v. 6.24 nAm List mean, [3.43 nAm sd]). We also identified two significant clusters in the vmPFC (185–312 ms,  $p = 0.043$ ; 10,000 permutations; 520–630 ms,  $p = 0.039$ ; 10,000 permutations). Again, localized activity during each cluster was greater during phrases compared to lists (cluster 1: 8.10 nAm Phrase mean, [3.59 nAm sd] v. 6.62 nAm List mean, [3.38 nAm sd]; cluster 2: 7.56 nAm Phrase mean, [4.84 nAm sd] v. 5.06 nAm List mean, [3.11 nAm sd]). Neither ROI showed any significant increases for the list

productions nor any significant effects during the view task (all clusters  $p > 0.8$ , 10,000 permutations).

In the LIFG, we observed a significant activity cluster (129–239 ms,  $p = 0.037$ ; 10,000 permutations) with significantly greater activity during lists compared to phrases (2.86 nAm Phrase mean, [0.95 nAm sd] v. 3.55 nAm List mean, [0.98 nAm sd]). We observed no increased LIFG activity for phrase productions over list productions nor any reliable effects between the two stimulus types in the view task (all clusters  $p > 0.8$ ; 10,000 permutations).

Given the sensitivity of anterior MEG activity to eye movements, we used our eye-tracking recordings to assess whether our LATL and vmPFC effects could have possibly resulted from increased eye-movements during the phrasal productions. Due to recording problems, eye-tracking data were reliably recorded only for 12 participants and thus this subset was analyzed. For this subset, the increases for phrase over list productions were replicated both in the LATL (cluster 1: 246–413 ms;  $p = 0.033$ ; 8.65 nAm Phrase mean [3.45 nAm sd] v. 5.97 nAm List mean, [2.49 nAm sd]; cluster 2: 424–700 ms;  $p = 0.020$ ; 8.16 nAm Phrase mean [4.33 nAm sd] v. 5.48 nAm List mean, [2.27 nAm sd]) and the vmPFC (cluster 1: 183–430 ms;  $p < 0.001$ ; 7.28 nAm Phrase mean [1.82 nAm sd] v. 5.78 nAm List mean, [2.59 nAm sd]; cluster 2: 458–632 ms;  $p = 0.022$ ; 7.52 nAm Phrase mean [4.38 nAm sd] v. 5.08 nAm List mean, [3.21 nAm sd]) while no differences were observed in the View-only control task (all clusters  $p > 0.35$ ; 10,000 permutations). Importantly, eye movements did not pattern with these ROI results: on average, there were more saccades for the list than the phrase productions within the 200–700 ms interval during which the reliable MEG effects occurred ( $t(11) = 3.9916$ ,  $p = 0.0021$ ). This is unsurprising given that the pictures eliciting the list productions required attention to the left and right parts of the stimulus, whereas the phrasal stimuli were a single object. During the view-only control task, the means patterned oppositely, with phrase pictures eliciting more eye-movements than list pictures, but this difference was not significant ( $p = 0.1$ ). Thus the eye-movement pattern clearly did not correlate with the MEG results. Continuous measures of pupil size as well as horizontal and vertical eye movements were additionally analyzed but showed no differences between conditions either.

In the ROI analysis reported above, we did not exclude trials with response times faster than 700 ms (the end of our epoch), as we expected such trials to be rejected by our artifact rejection routine. Further analysis of the response time data, however, revealed that on average, each subject's analyzed MEG data contained 24.4 trials [33.1 std.] with RTs faster than 700 ms. Interestingly, this implies that these productions did not create substantial motion artifacts. Although excluding these trials resulted in reduced power, we reran our analyses with somewhat looser parameters for multiple comparisons correction to assess whether the results of the main analysis would replicate for this subset of the data. [Supplementary material Fig. S1](#) plots the results of this analysis, which qualitatively replicates the pattern of LATL and vmPFC findings of the main analysis. However, in this analysis the LIFG effects did not replicate, suggesting that the LIFG increase during



**Fig. 2.** Experiment 1 ROI results. ROI results for the (A) LATL, (B) vmPFC, and (C) LIFG ROIs for both tasks, averaged across subjects. On the left, the locations of the ROIs are indicated in yellow. On the waveform plots, the shaded regions indicate significantly greater activity during phrase productions while the boxed region indicates significantly increased activity during list productions. Significance was determined using a non-parametric, permutation test (Maris & Oostenveld, 2007) performed over the entire epoch (10,000 permutations). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

list production depended in part on the trials with RTs faster than 700 ms. However, given the earliness of the LIFG effect in the main analysis, at around  $\sim 200$  ms, it is unlikely that this effect was driven directly by motion artifacts since on average, only 1% of the trials elicited production times faster than 500 ms.

### 2.2.3. Full-brain results

Our full-brain analyses (Fig. 3) conform closely to the ROI results and reveal clear combinatorial effects visible in both the vmPFC (100–200 ms) and LATL (300–600 ms). The lack of visible LIFG activity for list productions reflects the relatively weak and diffuse nature – in both space and time – of this effect, which becomes clarified by aggregating over both dimensions in the ROI analysis. Additionally, we observed early increased activity in the superior parietal lobule (SPL) during phrase productions and later increased activity in the right inferior temporal cortex (rIT) during list productions. While the exact functional significance of these effects is unclear, previous studies have implicated the SPL in sentence production (Haller et al., 2005) and the rIT in color discrimination (Zeki & Marini, 1998).

### 2.3. Discussion

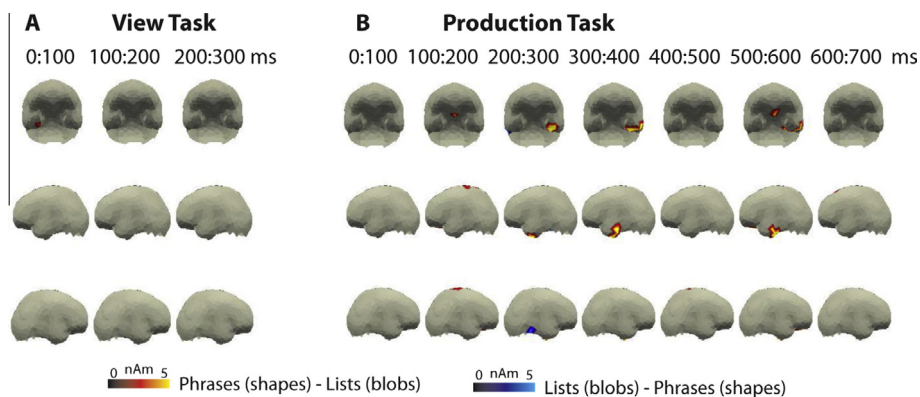
In this experiment, we aimed to characterize the spatio-temporal distribution of basic combinatorial brain activity during a simple two-word language production task. Our analyses focused on testing whether the LATL and the vmPFC, previously implicated for the combinatorial processing of similar stimuli in comprehension, would show evidence of combinatorial effects in production as well, thus lending support to models where comprehension and production share a compositional system. Consistent with this hypothesis, both regions exhibited a robust and temporally sustained activity increase for the phrasal stimuli compared to the list stimuli. Temporally, the vmPFC effect onset earlier than the LATL effect, which is the reverse of

the order previously observed in comprehension (Bemis & Pykkänen, 2011).

Although our stimulus choices achieved a manipulation of composition within expressions that were both natural and well-matched in various lexical statistics, an obvious difference between the phrase and list stimuli was that only the phrases required the production of a noun. Given that our comprehension studies on adjective–noun phrases have used noun–noun lists as the non-combinatorial control condition (Bemis & Pykkänen, 2011; Bemis & Pykkänen, 2012b), it is unlikely that the LATL and vmPFC effects observed here reflected increased responses to nouns as opposed to adjectives. Nevertheless, to formally rule out this hypothesis, we conducted a second experiment assessing whether similar effects would be observed if the control condition also involved a noun.

### 3. Experiment 2

The goal of Experiment 2 was to test whether the activity increases in Experiment 1 for phrasal ‘red cup’ productions over ‘red, blue’ list productions were due to combinatorial processing or to the presence of a noun in the phrasal productions. Although our first experiment avoided one word productions as the control condition due to the inevitably faster reaction times for planning the production of a single word as opposed to two, the early onset of the effects obtained in Experiment 1 made one-word productions an attractive option for Experiment 2: even if the control condition elicited somewhat faster reaction times, the majority of the significant effects observed in Experiment 1 should still be observable. Also, one-word productions allowed us to use simpler picture stimuli than, say, two-word noun lists (boat, cup) – previously used in our comprehension studies (Bemis & Pykkänen, 2011, 2012b) – which would have required a picture with two objects displayed simultaneously, worryingly more complex than the single colored object used to elicit the adjective–noun combinations. Thus one-word productions were chosen as the control stimuli of



**Fig. 3.** Experiment 1 full brain results. The results of the full-brain analyses are shown 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 20 ms and 15 contiguous sources). Red denotes increased activity during phrase productions (or colored shape viewing). Blue represents increased activity during list productions (or colored blob viewing). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Experiment 2 as this allowed us to (i) keep all productions highly natural and (ii) employ the same physical stimulus in each condition, eliminating the need for a separate View-only control task. Specifically, we had participants name the same colored objects either with just a noun ('tree'), just a color-adjective ('red') or with a noun-adjective combination describing both the color and the object ('red tree') (Fig. 4A–C). If the results of Experiment 1 were due to composition, the phrasal productions should elicit higher LATL and vmPFC activity than both the Noun-only and Adjective-only conditions. In contrast, if the results were due to the presence of a noun in the combinatory but not in the list-productions, then the Phrase vs. Noun-only contrast should not replicate the findings of Experiment 1, rather, the Phrase and Noun-only conditions should elicit increased activation as compared to the Adjective-only condition. This experiment was conducted in our MEG facility on the NYU Abu Dhabi campus, in substantially less noisy conditions than Experiment 1 (conducted in lower Manhattan).

### 3.1. Materials and methods

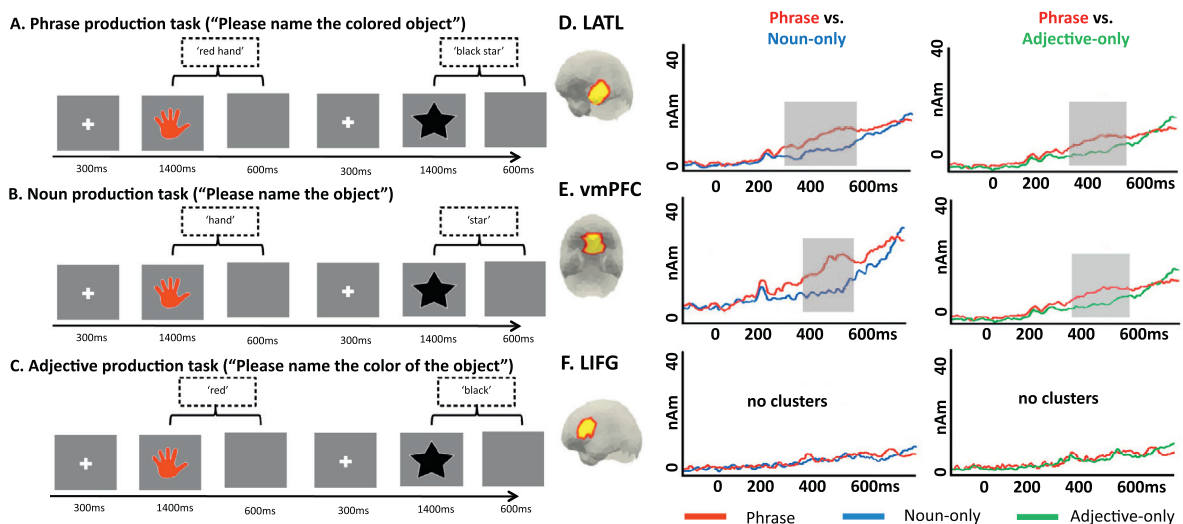
#### 3.1.1. Participants

14 right-handed, native English speakers participated in the experiment (2 female, 12 male, 25.5 years average 7.7 sd). All were neurologically intact, with normal or corrected-to-normal vision and all provided informed written consent.

#### 3.1.2. Stimuli

The experiment consisted of 294 trials divided into three conditions of 98 trials each. On each trial, participants were presented with a colored picture (e.g., a red

tree) and, depending on condition, were asked to either name aloud the noun describing the object in the picture (Noun-only condition: 'tree'), the adjective describing the color of the object (Adjective-only condition: 'red'), or to describe the colored object with an adjective–noun combination (Phrase condition: 'red tree') (Fig. 4A–C). The stimuli were the same as in the production task of Experiment 1 and were identical for each of the three conditions, thus eliminating the need for a View-only control task. Each experimental condition was further divided in 3 blocks of 33, 33 and 32 items respectively; therefore, the experiment consisted of a total of 9 blocks. The blocks always cycled through the order 'Noun-only - Phrase - Adjective-only' and participants were allowed to rest between blocks. Since the main goal of this experiment was to test whether the activity increases in Experiment 1 for 'red cup' type productions over 'red, blue' productions were due to combinatory processing or to the presence of a noun in the phrasal productions, we sequenced the stimuli such that the Noun-only productions were the productions with the least amount of lexical repetition of the three conditions, thus biasing that condition against activity decreases as compared to the phrasal productions. The Adjective-only condition was included solely to safeguard the experiment against a null result, should the findings of Experiment 1 indeed be due to the noun vs. adjective contrast. The fact that the Adjective-only condition was ordered last in the sequence of mini-blocks effectively gave it the strongest chance to exhibit priming-related activity decreases, biasing the experiment in favor of the result pattern where both conditions involving nouns would elicit increased activation as compared to the Adjective-only conditions. Thus if the composition-related activity pattern was observed despite this biasing, this would constitute



**Fig. 4.** Experiment 2 design and ROI results. The same colored objects were presented in each condition, but depending on the task instruction, participants described them either by using an adjective–noun phrase (A), a single noun (B) or a single adjective (C). ROI activity elicited by the phrasal productions was compared in pairwise cluster-based permutation tests to both one-word controls in the three ROIs employed in Experiment 1 (D–F). Both the LATL and the vmPFC showed robust and sustained activity increases for Phrases over Nouns, ruling out a syntactic category-based explanation of the corresponding effects in Experiment 1. The comparison between Phrases and Adjectives patterned similarly (shaded boxes indicate corrected  $p < 0.05$ ). No clusters were observed in the LIFG in either comparison. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



very strong evidence that the relevant activity pattern in fact reflects composition and not processing related to lexical or syntactic word class.

All pictures were presented foveally using Presentation (Neurobehavioral System Inc., California, USA) and subtended in a range from 1.41° to 2.77° height and 2.01° to 3.21° width on a screen ~85 cm from the subject.

### 3.1.3. Procedure

Before recording, each subject's head shape was digitized using a Polhemus dual source handheld FastSCAN laser scanner (Polhemus, VT, USA). MEG data were collected in the Neuroscience of Language Lab in NYU Abu Dhabi using a whole-head 208 channel axial gradiometer system (Kanazawa Institute of Technology, Kanazawa, Japan) as subjects lay in a dimly lit, magnetically shielded room. Vocal responses were captured with an MEG compatible microphone (Shure PG 81, Shure Europe GmbH).

In all conditions, trials began with a fixation cross (300 ms), followed by the presentation of the stimuli. Unlike in Experiment 1, where the picture remained onscreen until speech onset began, stimuli in Experiment 2 remained onscreen for a fixed interval of 1400 ms which was followed by a 600 ms blank screen (Fig. 4). This modification was done to simplify the trial structure somewhat; also, the disappearance of the picture upon voice-onset was found to be potentially distracting as we piloted the script. Keeping the stimulus on during articulation also avoided the blinks that subjects tended to produce at the offset of the stimuli (see above). However, the overall time provided for subjects to respond was equal to Experiment 1 (mean RT for production in Experiment 1 = 829 ms + 1200 ms provided to respond = 2029 ms, in Experiment 2 participants were provided with 1400 ms of picture onscreen + 600 of blank = 2000 ms). The entire recording lasted ~30 min.

### 3.1.4. Data acquisition and preprocessing

MEG data were recorded at 1000 Hz (200 Hz low-pass filter) and epoched from 200 ms before to 700 ms after picture onset. For artifact rejection, a cut-off of 3000 fT was used for all participants except for two, one of which showed a smaller and the other a larger general amplitude range from the others; cut-offs of 2500 fT and 3500 fT were used for these subjects instead, respectively. As in Experiment 1, we also rejected trials containing blinks identified by manual inspection of the MEG data. Altogether, this resulted in the exclusion of 18.7% of the trials (13.4% sd), leaving 244.5 trials on average per subject (39.51 sd). Thus the amount of rejected trials was considerably reduced in comparison to Experiment 1, presumably due to the better noise-conditions and the simpler trial structure, eliciting fewer blinks. Although the noise conditions in our Abu Dhabi lab did not require a high pass filter (contrary to our New York site), a 1 Hz high pass filter was applied to keep the analysis parallel to Experiment 1. However, we also report an analysis without this filter (Fig. S2), given that ideally high pass filtering should be kept to a minimum (e.g., Acunzo, MacKenzie, & van Rossum, 2012). Pre-processed data were then averaged for each condition and subject, using the interval of -200–0 ms for baseline

correction. Averages were low-pass filtered at 40 Hz, as in Experiment 1.

### 3.1.5. Statistical analysis

L2 Minimum norm estimates of source activity were created for each average using BESA 5.3 (MEGIS Software GmbH). The primary analysis focused on the same regions of interest (the LATL, vmPFC, and LIFG) and used the same type of ROI analyses as Experiment 1. Cluster-based permutation tests were used to compare the activity elicited by the phrasal productions with that associated with the two one-word conditions in one-tailed pair-wise *t*-tests (testing only for increases for Phrases compared to the one-word controls). The same cluster-selection criteria as in Experiment 1 were utilized. The permutation tests were conducted over the interval 0–500 ms post-picture onset, aimed at excluding articulation related motion artifacts (voice onset time was in the 700–900 ms range, see below). Since Experiment 2 was conducted strictly to address the syntactic category related explanation of the LATL and vmPFC effects obtained in Experiment 2, no explorative full brain analyses were conducted.

## 3.2. Results

### 3.2.1. Behavioral results

A one-way ANOVA on response time (RT) data elicited by the three experimental conditions revealed a significant effect of condition ( $F(2, 39) = 586.5, p < 0.001$ ). As expected, planned two-tailed pair-wise *t*-tests showed this to be driven by slower RTs for the Phrase productions (845 ms mean [89 ms sd]) than for the Noun-only (773 ms mean [80 ms sd],  $t(13) = 29.024, p < 0.001$ ) or Adjective-only productions (749 ms mean [90 ms sd],  $t(13) = 30.725, p < 0.001$ ), presumably reflecting the longer planning time required for two-word than for one-word productions. Also, the Noun-only productions were slightly (24 ms) but reliably slower than the Adjective-only productions ( $t(13) = 9.345, p < 0.001$ ). As in Experiment 1, during the recording sessions it was clear that accuracy in the production task was at ceiling, therefore, we mirrored our procedure in Experiment 1 and forewent analyses of accuracy and rejections of incorrect trials from the MEG data.

### 3.2.2. ROI results

Fig. 4D–F plot ROI activity in the three ROIs of Experiment 1 for pairwise comparisons between the Phrases and one-word control conditions. Activity increases are clearly observable for Phrases compared to both of the control conditions in both the LATL and the vmPFC, but clearly not in the LIFG, consistent with Experiment 1. The general waveform morphology and effect timecourse is somewhat more sustained than in Experiment 1. Indeed the statistical analysis of Experiment 2 revealed long sustained clusters as opposed to the seemingly biphasic pattern of Experiment 2 – this does not, however, rule out the possibility of two processing stages; instead, the two (or more) stages could be temporally contiguous or overlapping.

Cluster-based permutation tests were conducted in a parallel fashion to Experiment 1 and revealed a strong replication pattern for the main comparison of interest, i.e.,

the Phrase vs. Noun-only productions. In the LATL and the vmPFC, the Phrase productions elicited reliably increased activity compared to the Noun-only productions in long sustained clusters at 175–475 ms for the LATL ( $p = 0.0157$ , Fig. 4D) and at 285–500 ms for the vmPFC ( $p = 0.0165$ , Fig. 4E). These findings strongly rule out the hypothesis that the LATL and vmPFC effects in Experiment 1 were due to the presence of a noun in the Phrase but not in the List conditions: here, both conditions involved a noun and nevertheless the Phrases elicited higher activation in the same regions. Temporally, although the onset of the reliable vmPFC cluster was now later than the onset of the LATL cluster, contrary to Experiment 1, in the supplementary analysis where the 1 Hz high pass filter was removed (Fig. S2), the onsets of the two clusters were almost identical: 178 ms (vmPFC) vs. 182 ms (LATL). Thus, consistent with Experiment 1, these results conform to the generalization that vmPFC effects are not later than LATL effects in production.

The pattern observed for the Phrase vs. Adjective-only conditions was similar to that observed for the Phrase vs. Noun-only comparison. Again, the waveforms started separating in both regions around 200 ms, although now the reliable clusters onset somewhat later, at 266 ms for the LATL (266–500 ms,  $p = 0.0258$ ) and at 271 ms for the vmPFC (271–490 ms,  $p = 0.0136$ ). When the high-pass filter was removed, the LATL effect did however become significant by 205 ms while the vmPFC effect became marginal (Fig. S2). On the basis of these combined results, it can be concluded that  $\sim 200$  ms is the most consistent onset of the combinatorial effects – though the details may be sensitive to filter settings – and that the LATL effect is more robust and more straightforwardly replicable than the vmPFC effect. No clusters were identified in the LIFG for either pairwise comparison (Fig. 4F).

### 3.3. Discussion

Experiment 2 was a targeted investigation aimed at ruling out a syntactic category based explanation of the findings of Experiment 1, already unlikely on the basis of prior findings in comprehension (Bemis & Pylkkänen, 2011, 2013). If the LATL and vmPFC increases for Phrases over Lists in Experiment 1 had been due to the presence of a noun in the phrasal (red tree) but not in the list (red, blue) productions, including a noun as the control condition should have eliminated this effect. Instead, the LATL and vmPFC effects were robustly reproduced when the control condition was a noun, suggesting that this pattern of effects is related to composition and not the presence of a nominal category.

To safeguard the experiment against a possible null result, one-word adjective productions were also included in the design. The one-word adjectives also replicated the general pattern of Experiment 1, although not quite as cleanly, as the vmPFC effect became only a trend in our supplementary analysis without a high pass filter (Fig. S2). It should be noted though that the vmPFC effect was also more fragile in the Experiment 1, where it became somewhat fractionated and therefore marginal in the permutation test when the data were cleaned of

all RTs faster than 700 ms (Fig. S1). The same has been true for our comprehension studies using similar stimuli: although a clear vmPFC combinatorial effect was observed in our original basic composition study (Bemis & Pylkkänen, 2011), it was only weakly replicated in a follow-up study using both visual and auditory stimuli (Bemis & Pylkkänen, 2013). The overall pattern of these findings and our prior studies showing vmPFC effects of semantic complexity (Brennan & Pylkkänen, 2008, 2010; Pylkkänen & McElree, 2007; Pylkkänen, Martin, et al., 2009; Pylkkänen, Oliveri, et al., 2009) suggests that the activity localizing to the vmPFC is sensitive to semantic composition but as a prefrontal region, this effect is easily disrupted, perhaps due to the many sources of artifacts that can affect activity localizing close to the face (eyes, mouth, etc.).

## 4. General discussion

### 4.1. Combinatorial LATL and vmPFC effects

Using MEG, we investigated the spatio-temporal unfolding of neural activity generated during the production of simple adjective–noun phrases ('red tree') and observed significantly greater activity localized to both the LATL and vmPFC compared to noncombinatorial control stimuli. In Experiment 1 we controlled for the number of produced words and a variety of lexical statistics by using adjective–adjective lists as the control productions ('red, blue'), while in Experiment 2 we had subjects produce the individual lexical items of the combinatorial phrases as the control conditions (Noun-only 'tree'; Adjective-only 'red'). Crucially, in Experiment 2, the picture to be named was the same for all conditions, eliminating any concern of perceptually driven effects. In both Experiments, the combinatorial effects in the LATL were more robustly reproducible across comparisons and analyses than those in the vmPFC, suggesting the latter effect is somewhat more fragile.

This pattern of effects spatially parallels our previous findings on the comprehension of similar phrases (Bemis & Pylkkänen, 2011), supporting a shared brain basis for composition during production and comprehension. With MEG we were able to characterize the detailed time-evolution of both effects, which in comprehension occurred at  $\sim 250$  ms for the LATL and at  $\sim 400$  ms for the vmPFC. During the current production tasks however, the vmPFC effect no longer followed the LATL effect but rather had a similar (Exp. 2) or an even earlier onset (Exp. 1). The fact that vmPFC effects are late in comprehension but early in production is broadly consistent with the notion that production is to some degree the reverse of comprehension. Although the exact onsets of the reliable LATL and vmPFC clusters varied somewhat depending on stimulus and filtering details, the wave separations between the phrasal and control conditions systematically started around  $\sim 200$  ms. This timing fits perfectly to the prediction arising from the combination of psycholinguistic models, which hypothesize that combinatorial information becomes available during lexical access (Ferreira, 2000; Ferreira &

Slevc, 2007; Levelt, 1989), and single word electrophysiological production studies, which place lexical access concurrent with the initiation of our observed combinatorial effects (Indefrey & Levelt, 2004; Strijkers & Costa, 2011). The subsequent waveform morphologies showed a rather sustained profile in both the LATL and vmPFC, with Experiment 1 providing some evidence for a potentially biphasic activity pattern in the vmPFC.

Psycholinguistic models (Frazier, 1987; Levelt, 1989) hypothesize that comprehension and production rely on functionally similar combinatorial operations, such as the establishment of syntactic and semantic relationships between constituent elements. To date, however, evidence for shared processing derives primarily from structural priming effects (Pickering & Ferreira, 2008) and hemodynamic measures of shared activity (Awad et al., 2007; Menenti et al., 2011) and is thus unable to determine temporal relationships between the two. The current study reveals long stretches of combinatory activity in two brain regions previously implicated in composition during comprehension. Undoubtedly, these sustained activations must reflect multiple processing steps and thus a pivotal goal for future work will be to characterize the internal computational organization of these activities. Although our current study does not go beyond demonstrating that regionally, combinatory effects localize similarly in production and comprehension, and temporally, the onsets of the two earliest effects largely parallel, in what follows we outline some possibilities regarding the functional roles of the observed effects, in light of extant psycholinguistic and cognitive neuroscience models.

Given that many models place combinatorial syntactic operations before semantic operations during comprehension (Frazier, 1987, 2002) and after semantic operations during production (Indefrey & Levelt, 2004; Levelt, 1989), one possibility consistent with the present findings and the corresponding comprehension results (Bemis & Pykkänen, 2011) is that the LATL effect reflects syntactic and the vmPFC effect semantic combinatory operations. While this proposal is reasonably unproblematic for the vmPFC given its systematic sensitivity to semantic complexity in syntactically parallel expressions (Brennan & Pykkänen, 2008, 2010; Pykkänen & McElree, 2007; Pykkänen, Martin, et al., 2009; Pykkänen, Oliveri, et al., 2009), the evidence on the LATL as a syntactic region is quite conflicted. Although several studies have found that combinatory effects in the LATL can be obtained even in the absence of lexical semantics (Friederici et al., 2000; Humphries et al., 2006), this has not uniformly been true (Pallier, Devauchelle, & Dehaene, 2011) and the LATL has also shown sensitivity to factors such as the conceptual congruency or plausibility of a sentence (Humphries et al., 2006; Mazoyer et al., 1993; Vandenberghe, Nobre, & Price, 2002). Further, a sizeable separate literature on semantic memory has proposed the LATL as a feature binding hub that operates across a distributed semantic network (Lambon Ralph, Pobric, & Jefferies, 2009; Lambon Ralph, Sage, Jones, & Mayberry, 2010; Patterson, Nestor, & Rogers, 2007). In fact, single word conceptual specificity has recently been shown to reliably affect the size of the LATL combinatory effect within syntactically parallel

phrases in MEG, supporting a semantic as opposed to a syntactic account of the LATL and demonstrating that LATL effects of composition and conceptual specificity are likely to be related (Westerlund & Pykkänen, 2014). In sum, although the evidence is somewhat mixed, it is unlikely that the LATL is a semantically insensitive phrase structure builder in any straightforward sense.

A more nuanced possibility could perhaps employ insights from Binder and Desai's (2011) theory of 'embodied abstraction,' in which different levels of conceptual representation involve different degrees of abstraction and sensory-motor/affective simulation. Under this type of an account, combinatory operations could conceivably take as input conceptual representations at different levels of abstraction, the LATL perhaps corresponding to a more abstract, but possibly still semantic, level of composition. The vmPFC, in contrast, would operate on more "fleshed out," later representations (Binder & Desai, 2011). Given that during picture naming, the full conceptual representation is immediately constructable from the picture, one would expect early vmPFC activity during production, feeding perhaps in parallel the construction of more abstract representations, built by hypothesis in the LATL. Though simply a speculation at this point, this conceptualization of the data would also conform well with recent theories of the vmPFC as a high level integrative site involved in the generation of "affective meaning" across a variety of domains (Roy et al., 2012).

An important aspect of the current findings is that in addition to the altered relative ordering of the onsets of the LATL and vmPFC effects as compared to their temporal profile in comprehension, these early effects were either followed by a second stage of effects in the same sources (Exp. 1) or fed into long sustained effects lasting until about 500 ms (Exp. 2). Depending on many specifics regarding data preprocessing and between-subjects variability in effect latencies, separate stages of effects can easily blend into one, and thus we do not believe much should be made about the biphasic vs. sustained effect profiles of Experiments 1 vs. 2. Crucially though, these effects are much longer lasting than what we have observed in comprehension (Bemis & Pykkänen, 2011, 2012b). This finding resonates with theories in which production incorporates comprehension processes in the form of speakers monitoring their own (planned or already uttered) utterances (Levelt, 1989) – in other words, after constructing the to-be-uttered utterance, the producer comprehends it internally. A recent model explicitly incorporating a full comprehension module into the production process is Pickering and Garrod's (2013), where during production, the "production implementer" first creates a to-be-produced representation by mapping higher representational levels to lower ones (from semantics to syntax to phonology), and this stage is immediately followed by the "comprehension implementer" which understands the just created [sem, syn, phon] triplet by processing the levels in the opposite order, i.e., from phonology to semantics. Although our study was in no way designed to test the specific predictions of this type of model, the temporal profile of our LATL and vmPFC effects is quite consistent with the gist of this account. Thus the current results should provide

a useful starting point for more targeted hypothesis testing on the computational details of the comprehension–production relationship.

#### 4.2. LIFG activity

Given the prominence of the LIFG in language production research, it was natural to include this ROI in our study, however we observed no significant LIFG increases for phrase productions as compared to lists. Previous effects in this region during production have primarily been observed during rather complex utterances such as sentences (Haller et al., 2005; Menenti et al., 2011) and narratives (Braun et al., 2001), and increased LIFG activity has been associated with several cognitive functions that might plausibly be involved during such productions, such as working memory (Chein, Fissell, Jacobs, & Fiez, 2002) and lexical/semantic selection (Heim, Eickhoff, Friederici, & Amunts, 2009; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997). Even previous results linking the LIFG to simpler productions (Indefrey et al., 2001) may reflect such non-combinatorial processing, e.g. selecting gender information (necessary for German) which recruits the LIFG during single-word production as well (Heim, Opitz, & Friederici, 2002).

Of these possibilities, our results are most consistent with proposals linking the LIFG to selection at the lexical-semantic level, as we did observe increased LIFG activity during list productions ('red, blue') compared to phrases in Experiment 1. However, it should be noted that this effect was fragile and not replicated in our supplementary analysis with a cleaner, but smaller subset of the trials (Fig. S1). Semantic interference during the production of two semantically similar words has previously been associated with increased LIFG activity (Heim et al., 2009; Schnur et al., 2009), and such interference also delays speech onset (Smith & Wheeldon, 2004), as we observed in our list condition. Further, electrophysiological semantic interference effects have been observed at the same time as the present LIFG effect (Aristei, Melinger, & Abdel Rahman, 2011; Maess, Friederici, Damian, Meyer, & Levelt, 2002). Thus, the present findings are broadly compatible with theories attributing a role for the LIFG in semantic selection (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997), although they do not speak to subtler questions regarding this type of theory, such as whether the LIFG reflects competition among alternatives or the resolution of competition (Grindrod, Bilenko, Myers, & Blumstein, 2008). The general selection-based hypothesis also straightforwardly predicts the completely null LIFG results of Experiment 2, where the control conditions only involved one word and thus elicited no interference.

#### 4.3. Conclusions

The present study provides the first directed investigation into the spatio-temporal dynamics of basic combinatorial neural mechanisms that subservise language production. Using MEG, we compared neural activity generated during the production of simple adjective–noun phrases ('red tree') with non-compositional list ('red, blue')

and one-word controls ('red', 'tree'). We identified early combinatorial effects beginning at ~180 ms, suggesting that combinatorial processing initiates with lexical access during production. Spatially, combinatorial activity localized to the LATL and vmPFC, coinciding with that observed during the comprehension of similar phrases (Bemis & Pyllkkänen, 2011). Further, while in comprehension, combinatorial activity in the LATL has preceded effects in the vmPFC, vmPFC effects now either preceded or occurred in parallel with LATL effects. Thus, although prior hemodynamic research has already suggested a shared neural architecture for production and comprehension, the present findings constitute the first direct neural link between psycholinguistic models of comprehension and production that posit functionally similar combinatorial processes operating in potentially reversed temporal order. Future work is now needed to identify the precise functional makeup of this combinatorial, bidirectional neural pathway.

#### Acknowledgements

This research was funded by the Whitehead Fellowship for Junior Faculty in Biomedical and Biological Sciences (LP), National Science Foundation Grant BCS-1221723 (LP) and Grant G1001 from the NYUAD Institute, New York University Abu Dhabi (LP). We thank Victor Ferreira for discussion of the results, Anna Bemis for her assistance in creating the figures and Paul Del Prato for help with the data analysis.

#### Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cognition.2014.07.001>.

#### References

- Acunzo, D. J., MacKenzie, G., & van Rossum, M. C. (2012). Systematic biases in early ERP and ERF components as a result of high-pass filtering. *Journal of Neuroscience Methods*, *30*, 212–218.
- Alario, F., Costa, A., & Caramazza, A. (2002). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes*, *17*, 299–319.
- Aristei, S., Melinger, A., & Abdel Rahman, R. (2011). Electrophysiological chronometry of semantic context effects in language production. *Journal of Cognitive Neuroscience*, *23*, 1567–1586.
- Awad, M., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. S. (2007). A common system for the comprehension and production of narrative speech. *The Journal of Neuroscience*, *27*, 11455–11464.
- Bemis, D. K., & Pyllkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *The Journal of Neuroscience*, *31*, 2801–2814.
- Bemis, D. K., & Pyllkkänen, L. (2012a). Combination across domains: An MEG investigation into the relationship between mathematical, pictorial, and linguistic processing. *Frontiers in Psychology*, *3*.
- Bemis, D., & Pyllkkänen, L. (2012b). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*. <http://dx.doi.org/10.1093/cercor/bhs179>.
- Bemis, D. K., & Pyllkkänen, L. (2013). Flexible composition: MEG evidence for the deployment of basic combinatorial linguistic mechanisms in response to task demands. *PLoS ONE*, *8*(9), e73949.
- Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536.



- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Braun, A. R., Guillemin, A., Hoesly, L., & Varga, M. (2001). The neural organization of discourse. *Brain*, *124*, 2028–2044.
- Brennan, J., & Pykkänen, L. (2008). Processing events: Behavioral and neuromagnetic correlates of aspectual coercion. *Brain and Language*, *106*, 132–143.
- Brennan, J., & Pykkänen, L. (2010). Processing psych verbs: Behavioral and MEG measures of two different types of semantic complexity. *Language and Cognitive Processes*, *25*, 777–807.
- Brennan, J., & Pykkänen, L. (2012). The time-course and spatial distribution of brain activity associated with sentence processing. *Neuroimage*, *60*(2), 1139–1148.
- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network. *Annals of the New York Academy of Sciences*, *1124*(1), 1–38.
- Chein, J. M., Fissell, K., Jacobs, S., & Fiez, J. A. (2002). Functional heterogeneity within Broca's area during verbal working memory. *Physiology & Behavior*, *77*, 635–639.
- Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, *99*, 205–235.
- Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J., et al. (2002). The speaking brain: A tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. *Journal of Neurolinguistics*, *15*, 59–90.
- Eulitz, C., Hauk, O., & Cohen, R. (2000). Electroencephalographic activity over temporal brain areas during phonological encoding in picture naming. *Clinical Neurophysiology*, *111*, 2088–2097.
- Ferreira, F. (2000). Syntax in language production: An approach using tree-adjointing grammars. In L. Wheelodon (Ed.), *Aspects of language production* (pp. 291–330). Hove, United Kingdom: Psychology Press.
- Ferreira, V. S., & Slevc, L. R. (2007). Grammatical encoding. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 453–470). Oxford: Oxford University Press.
- Frazier, L. (1987). Theories of sentence processing. In J. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. 291–307). MIT Press.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *TRENDS in Cognitive Sciences*, *6*, 78–84.
- Friederici, A. D., Meyer, M., & von Cramon, D. Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language*, *74*(2), 289–300.
- Grindrod, C. M., Bilenko, N. Y., Myers, E. B., & Blumstein, S. E. (2008). The role of the left inferior frontal gyrus in implicit semantic competition and selection: An event-related fMRI study. *Brain Research*, *1229*, 167.
- Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience*, *2*(10), 685–694.
- Habets, B., Jansma, B. M., & Münte, T. F. (2008). Neurophysiological correlates of linearization in language production. *BMC Neuroscience*, *9*, 77.
- Haller, S., Radue, E. W., Erb, M., Grodd, W., & Kircher, T. (2005). Overt sentence production in event-related fMRI. *Neuropsychologia*, *43*, 807–814.
- Heim, S., Eickhoff, S. B., Friederici, A. D., & Amunts, K. (2009). Left cytoarchitectonic area 44 supports selection in the mental lexicon during language production. *Brain Structure & Function*, *213*, 441–456.
- Heim, S., Opitz, B., & Friederici, A. D. (2002). Broca's area in the human brain is involved in the selection of grammatical gender for language production: Evidence from event-related functional magnetic resonance imaging. *Neuroscience Letters*, *328*, 101–104.
- Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, *18*(4), 665–679.
- Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Human Brain Mapping*, *26*(2), 128–138.
- Indefrey, P., Brown, C. M., Hellwig, F., Amunts, K., Herzog, H., & Hagoort, P. (2001). A neural correlate of syntactic encoding during speech production. *Proceedings of the National Academy of Sciences*, *98*, 5933–5936.
- Indefrey, P., Hellwig, F., Herzog, H., Seitz, R. J., & Hagoort, P. (2004). Neural responses to the production and comprehension of syntax in identical utterances. *Brain and Language*, *89*, 312–319.
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.
- Lambon Ralph, M. A., Pobric, G., & Jefferies, E. (2009). Conceptual knowledge is underpinned by the temporal pole bilaterally: Convergent evidence from rTMS. *Cerebral Cortex*, *19*(4), 832–838.
- Lambon Ralph, M. A., Sage, K., Jones, R. W., & Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences*, *107*(6), 2717–2722.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, *10*, 553–567.
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*, 203–208.
- Maess, B., Friederici, A. D., Damian, M., Meyer, A. S., & Levelt, W. J. M. (2002). Semantic category interference in overt picture naming: Sharpening current density localization by PCA. *Journal of Cognitive Neuroscience*, *14*, 455–462.
- Maguire, E. A., Frith, C. D., & Morris, R. G. M. (1999). The functional neuroanatomy of comprehension and memory: The importance of prior knowledge. *Brain*, *122*(10), 1839–1850.
- Malpass, D., & Meyer, A. S. (2010). The time course of name retrieval during multiple-object naming: Evidence from extrafoveal-on-foveal effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 523–537.
- Mar, R. A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia*, *42*(10), 1414–1434.
- Marek, A., Habets, B., Jansma, B. M., Nager, W., & Münte, T. F. (2007). Neural correlates of conceptualisation difficulty during the preparation of complex utterances. *Aphasiology*, *21*, 1147–1156.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.
- Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., et al. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, *5*(4), 467–479.
- Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science*, *22*, 1173–1182.
- Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, *35*, 477–496.
- Meyer, A. S., Ouellet, M., & Häcker, C. (2008). Parallel processing of objects in a naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 982–987.
- Nathaniel-James, D. A., & Frith, C. D. (2002). The role of the dorsolateral prefrontal cortex: Evidence from the effects of contextual constraint in a sentence completion task. *Neuroimage*, *16*, 1094–1102.
- Ojemann, J. G., Akbudak, E., Snyder, A. Z., McKinsty, R. C., Raichle, M. E., & Conturo, T. E. (1997). Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *Neuroimage*, *6*(3), 156–167.
- Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273–281.
- Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, *108*(6), 2522–2527.
- Palmer, E. D., Rosen, H. J., Ojemann, J. G., Buckner, R. L., Kelley, W. M., & Petersen, S. E. (2001). An event-related fMRI study of overt and covert word stem completion. *Neuroimage*, *14*, 182–193.
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347.
- Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*, 427–459.
- Potter, M. C., & Faulconer, B. A. (1979). Understanding noun phrases. *Journal of Verbal Learning and Verbal Behavior*, *18*, 509–521.
- Pykkänen, L., Brennan, J., & Bemis, D. K. (2011). Grounding the cognitive neuroscience of semantics in linguistic theory. *Language and Cognitive Processes*, *26*(9), 1317–1337.

- Pykkänen, L., Oliveri, B., & Smart, A. (2009a). Semantics vs. world knowledge in prefrontal cortex. *Language and Cognitive Processes*, 24, 1313–1334.
- Pykkänen, L., Martin, A. E., McElree, B., & Smart, A. (2009b). The anterior midline field: Coercion or decision making? *Brain and Language*, 108, 184–190.
- Pykkänen, L., & McElree, B. (2007). An MEG study of silent meaning. *Journal of Cognitive Neuroscience*, 19, 1905–1921.
- Rogalsky, C., & Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex*, 19(4), 786–796.
- Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, 16(3), 147–156.
- Salmelin, R. (2007). Clinical neurophysiology of language: The MEG approach. *Clinical Neurophysiology*, 118, 237–254.
- Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences*, 106, 322–327.
- Schriefers, H., De Ruiter, J. P., & Steigerwald, M. (1999). Parallelism in the production of noun phrases: Experiments and reaction time models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 702–720.
- Smith, M., & Wheeldon, L. (2004). Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 675–686.
- Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J. A., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex*, 19(7), 1493–1503. <http://dx.doi.org/10.1093/cercor/bhn187>.
- Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32), 14425–14430.
- Stowe, L. A., Broere, C. A., Paans, A. M., Wijers, A. A., Mulder, G., Vaalburg, W., et al. (1998). Localizing components of a complex task: Sentence processing and working memory. *NeuroReport*, 9(13), 2995–2999.
- Strijkers, K., & Costa, A. (2011). Riding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Frontiers in Psychology*, 2, 1–16.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, 94(26), 14792–14797.
- Vandenbergh, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*, 14(4), 550–560.
- Westerlund, M., & Pykkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia*, 57, 59–70.
- Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: Emergent features of word, sentence, and narrative comprehension. *NeuroImage*, 25(3), 1002–1015.
- Zeki, S., & Marini, L. (1998). Three cortical stages of colour processing in the human brain. *Brain*, 121, 1669–1685.
- Zimmermann, T. E. (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics*, 8, 255–290.