



## Research report

# The role of the IFG and pSTS in syntactic prediction: Evidence from a parametric study of hierarchical structure in fMRI



William Matchin <sup>a,\*</sup>, Christopher Hammerly <sup>b</sup> and Ellen Lau <sup>c</sup>

<sup>a</sup> Department of Linguistics, UC San Diego, United States

<sup>b</sup> Department of Linguistics, University of Massachusetts Amherst, United States

<sup>c</sup> Department of Linguistics, University of Maryland, United States

## ARTICLE INFO

## Article history:

Received 22 March 2016

Reviewed 28 June 2016

Revised 1 September 2016

Accepted 9 December 2016

Action editor Cynthia Thompson

Published online 18 December 2016

## Keywords:

Broca's area

Prediction

Language

Sentence processing

Syntax

## ABSTRACT

Sentences encode hierarchical structural relations among words. Several neuroimaging experiments aiming to localize combinatory operations responsible for creating this structure during sentence comprehension have contrasted short, simple phrases and sentences to unstructured controls. Some of these experiments have revealed activation in the left inferior frontal gyrus (IFG) and posterior superior temporal sulcus (pSTS), associating these regions with basic syntactic combination. However, the wide variability of these effects across studies raises questions about this interpretation. In an fMRI experiment, we provide support for an alternative hypothesis: these regions underlie top-down syntactic predictions that facilitate sentence processing but are not necessary for building syntactic structure. We presented stimuli with three levels of structure: unstructured lists, two-word phrases, and simple, short sentences; and two levels of content: natural stimuli with real words and stimuli with open-class items replaced with pseudowords (jabberwocky). While both the phrase and sentence conditions engaged syntactic combination, our experiment only encouraged syntactic prediction in the sentence condition. We found increased activity for both natural and jabberwocky sentences in the left IFG (*pars triangularis* and *pars orbitalis*) and pSTS relative to unstructured word lists and two-word phrases, but we did not find any such effects for two-word phrases relative to unstructured word lists in these areas. Our results are most consistent with the hypothesis that increased activity in IFG and pSTS for basic contrasts of structure reflects syntactic prediction. The *pars opercularis* of the IFG showed a response profile consistent with verbal working memory. We found incremental effects of structure in the anterior temporal lobe (ATL), and increased activation only for sentences in the angular gyrus (AG)/temporal–parietal junction (TPJ) – both regions showed these effects for stimuli with all real words. These findings support a role for the ATL in semantic combination and the AG/TPJ in thematic processing.

© 2016 Elsevier Ltd. All rights reserved.

\* Corresponding author. Department of Linguistics, UC San Diego, 9500 Gilman Drive #0108, La Jolla, CA 92093-0108, United States

E-mail address: [wmatchin@ucsd.edu](mailto:wmatchin@ucsd.edu) (W. Matchin).

<http://dx.doi.org/10.1016/j.cortex.2016.12.010>

0010-9452/© 2016 Elsevier Ltd. All rights reserved.

## 1. Introduction

Language does not consist merely of sequences of words, but rather intricate syntactic structures that determine semantic interpretation (Chomsky, 1965; Everaert, Huybregts, Chomsky, Berwick, & Bolhuis, 2015). Linguists have postulated a set of basic combinatory mechanisms that generate this structure (Chomsky, 1995; Jackendoff, 2003; Pollard & Sag, 1994), and psycholinguists have developed models of how this structure is built during comprehension and production (e.g., Frazier & Fodor, 1978; Lewis & Vasishth, 2005; Phillips, 1996). Neuroscience research on syntax has focused mostly on the left inferior frontal gyrus (IFG; Brodmann's areas 44, 45, and 47) (e.g., Caramazza & Zurif, 1976; Embick, Marantz, Miyashita, O'Neil, & Sakai, 2000; Friederici, Fiebach, Schlesewsky, Bornkessel, & von Cramon, 2006), the left anterior temporal lobe (ATL) (e.g., Brennan et al., 2012; Mazoyer et al., 1993), and the left posterior superior temporal sulcus (pSTS) (e.g., Friederici, Makuuchi, & Bahlmann, 2009; Stowe et al., 1998) as regions important for syntactic processing. However, many of these studies have used complex stimuli and tasks, raising the important question: do activations in these areas reflect basic combinatory mechanisms, or do they reflect additional mechanisms that contribute to sentence processing?

In light of this question, several recent studies have used simple and/or short phrases to identify brain regions involved in syntax (e.g., Bemis & Pykkänen, 2011; Pallier, Devauchelle, & Dehaene, 2011; Zaccarella, Meyer, Makuuchi, & Friederici, 2015). Some of these studies have associated activation in the IFG and the pSTS specifically with syntactic combination (Pallier et al., 2011; Tyler, Randall, & Stamatakis, 2008; Zaccarella et al., 2015). Additionally, other studies have shown that subregions of the IFG are sentence-selective (Fedorenko, Duncan, & Kanwisher, 2012; Rogalsky, Almeida, Sprouse, & Hickok, 2015). In other words, they activate to sentences and do not activate to non-linguistic tasks involving domain-general mechanisms that others have suggested underlie the function of the IFG in sentence processing (e.g., Novick, Trueswell, & Thompson-Schill, 2005; Rogalsky, Matchin, & Hickok, 2008). Altogether, these recent studies suggest more strongly that the IFG and pSTS support basic syntactic combination.

However, there remain important reasons to doubt this conclusion. First, a simple and important expectation for brain areas involved in basic syntactic combination is that these regions consistently activate for sentences during neuroimaging studies. Yet many studies have failed to find increased activation for sentences or phrases compared to word lists or baseline in the IFG (e.g., Bemis & Pykkänen, 2011; Friederici, Kotz, Scott, & Oleser, 2010; Humphries, Love, Swinney, & Hickok, 2005; Mazoyer et al., 1993; Stowe et al., 1998) or in the pSTS (Humphries, Binder, Medler, & Liebenthal, 2006; Rogalsky & Hickok, 2009). With respect to the neuropsychological literature, patients with Broca's aphasia, associated with damage to the left IFG (Fridriksson, Fillmore, Guo, & Rorden, 2015), can determine whether a sentence is structurally well formed or not (Linebarger, Schwartz, & Saffran, 1983; Wilson & Saygin, 2004; Wulfeck & Bates, 1991), a task that depends on syntactic combination.

Furthermore, damage to the left IFG is not consistently associated with basic sentence comprehension deficits that would be expected with deficits in syntactic combination (Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004; Magnusdottir et al., 2013; Thothathiri, Kimberg, & Schwartz, 2012). In fact, damage to the IFG alone does not result in chronic Broca's aphasia, but rather milder deficits in speech production (Mohr et al., 1978). Similar to the literature on language disorders caused by focal brain trauma, the primary progressive aphasia literature has associated degeneration of the IFG (and to a lesser extent, the pSTS) with problems of grammatical processing (e.g., comprehension deficits with complex sentences and difficulties with morphology, inflection, functional lexical items, and verb argument structure; Thompson & Mack, 2014; Wilson, Galantucci, Tartaglia, & Gorno-Tempini, 2012). However, the more subtle tests of grammatical knowledge using acceptability judgments have not been performed to determine whether these patients have a loss of grammatical knowledge, or whether they show a similar pattern to agrammatism, with largely intact grammatical knowledge. The situation is also equivocal in the pSTS – damage to this region is associated with sentence comprehension deficits in some studies (Dronkers et al., 2004; Wilson et al., 2010) but not in others (Thothathiri et al., 2012). These data leave us with a paradox – the IFG and pSTS sometimes show basic effects of structure and sometimes do not, and damage to these regions, particularly the IFG, does not seem to impair basic structure-building processes.

We believe that this quandary is resolved by the hypothesis that increased IFG and pSTS activity for simple structural contrasts reflects top-down processes involving structure and syntactic features that facilitate sentence processing but are not necessary for building basic structure. In particular, we suggest *structural prediction*. That is, we posit that the IFG and pSTS together encode top-down predictions of phrasal nodes (e.g., NP, VP), syntactic features (e.g., agreement features like number, tense features), and perhaps syntactic information on specific lexical items or lemmas such as gender. Because top-down prediction facilitates sentence processing, particularly when input is rapid or incomplete, it is likely to be engaged by many sentence comprehension experiments and may account for some of the language deficits in patients with damage to these areas. The IFG and pSTS sub-areas that support structural prediction may also be language-specific (although we remain agnostic on this point). Critically, however, we posit that these predictive mechanisms are not necessary for building basic syntactic structure. Therefore these regions might not always activate during sentence comprehension, and damage will not necessarily result in sentence comprehension deficits.

Previous work provides good evidence that speakers make structural predictions in a variety of syntactic contexts. For instance, in filler-gap dependencies (e.g., in a sentence like *Diogo reviewed the paper<sub>1</sub> that Jon submitted \_\_<sub>1</sub>*, “the paper” is the filler, and “\_\_” stands for the gap), a classic finding is that processing is slowed when expectations of gaps are violated (the “filled-gap effect” – Crain & Fodor, 1985; Stowe, 1986). The early brain negativity elicited in electroencephalography (EEG) studies for syntactic violations, the ELAN (Friederici, 2002; Neville, Nicol, Barss, Forster, & Garrett, 1991), is attenuated

in ellipsis configurations when syntactic predictions are weaker (Lau, Stroud, Plesch, & Phillips, 2006). Overall, prediction effects have been observed across a variety of different syntactic constructions in many studies of on-line sentence processing (Konieczny, 2000; Levy & Keller, 2013; Omaki et al., 2015; Staub & Clifton, 2006; Yoshida, Dickey, & Sturt, 2013).

Although only a few neuroimaging studies have specifically tested structural prediction, four fMRI studies have shown that syntactic and/or lexical-semantic predictions in sentence processing are associated increased activity in the IFG. Santi and Grodzinsky (2012) showed that sentences with filler-gap dependencies activate the anterior IFG (*pars triangularis*) compared to sentences that do not, but only when these gaps are predicted. Increased activation is also observed in the anterior IFG when predictions of gaps or the antecedent to a pronoun have to be maintained over longer distances of intervening material (Matchin, Sprouse, & Hickok, 2014; Santi & Grodzinsky, 2007). Finally, Bonhage, Mueller, Friederici, and Fiebach (2015) showed in an fMRI experiment in German that predictions of the syntactic category of the final word of jaberwocky sentences (sentences with real words replaced by pseudowords) resulted in increased activation in left IFG and pSTS compared to unstructured lists where no such prediction was possible. Similarly, an ERP/neuropsychology study by Jakuszeit, Kotz, and Hasting (2013) used the ELAN response as a marker of syntactic prediction and examined this response to word category and agreement violations in patients with left IFG damage (without language disorder). They showed a stark and selective difference for subject-verb agreement: normal subjects showed a robust ELAN for agreement violations, while the patients showed no hint of an ELAN, although both groups showed sensitivity to these violations in a later time window. Consistent with the fMRI work, the authors interpreted these results as supporting the role of the IFG in syntactic predictions.

### 1.1. The present study

The structural prediction hypothesis is able to explain the results of neuroimaging experiments in which structured stimuli do not elicit IFG or pSTS activity because we assume that generating top-down predictions is costly – unless the task or materials encourage these mechanisms, subjects might not use them. In the fMRI experiment reported here, we provide novel support for this hypothesis by finding distinct effects in these regions for two different structured conditions in the same group of subjects. We used a parametric design similar to Pallier et al. (2011) with three levels of structure: unstructured word lists, sequences of minimal two-word phrases, and simple six-word sentences. Our sentences and phrases were syntactically canonical, simple, and short – this ensured that any activation we observed for sentences was not due to the complexity or difficulty of our materials. We also manipulated whether our stimuli had real content words (natural) or not (jabberwocky) – this allowed us to attempt to disentangle effects of syntax and semantics (e.g., Mazoyer et al., 1993; Pallier et al., 2011), although it remains likely that certain aspects of semantics remained in the jabberwocky materials, such as an abstract event structure (i.e., theta roles, or who did what to whom; Chomsky, 1982) in the

jabberwocky sentences. Our task required subjects to determine whether a probe word was present in the preceding stimulus, which required equal attention to all stimuli across our conditions (we collected behavioral data in the scanner to verify that subjects attended to all conditions). We used a block design that always presented stimuli in multiple six-word sequences, with the probe memory task occurring randomly after occasional sequences.

Both phrases and sentences involve basic syntactic combination, but we only expected the sentence condition to involve structural prediction given the nature of sentence processing from the psycholinguistic literature. Work stretching back decades has shown that hierarchical syntactic structure and semantic interpretation of sentences facilitates word recall in tasks similar to ours e.g., Brener, 1940; Miller, Heise, & Lichten, 1951; Miller & Isard, 1963, Marks & Miller, 1964). Essentially, subjects appear to approach a word recall task in sentence context by building the structure and interpreting the meaning of the sentence, and using these representations to guide their responses to the word recall task. If subjects were to use the whole structure and sentence meaning, but did not build the structure predictively, they would have to wait until the very end of the sentence to complete the structure and then interpret the meaning of the sentence. However, in our experiment there was little time in-between the end of the sentence and the deadline to make a response by the end of the probe, and predicting the structure (thereby also allowing prediction of the sentence meaning) would allow more preparation to respond to the probe and reduce the burden of remembering the words separately. For these reasons we expected that the predictive nature of sentence processing identified in the psycholinguistics literature reviewed above would even more strongly encourage subjects to generate structural predictions in our study. By our hypothesis, this should result in increased activity in the IFG and pSTS relative to the list conditions.

While we think that people can in principle generate syntactic and semantic predictions in local contexts, we expected that subjects would *not* use top-down structural prediction in our two-word phrase condition for two reasons. In the two-word phrase condition the final word of each phrase arrives immediately after the first. If top-down structural predictions are costly and most useful for projecting hierarchical structure beyond immediately adjacent words, subjects might not predict structure for two-word phrases (i.e., the syntactic category for the immediately following word). Even if predictions *were* generated in the phrase condition, they would not have to be maintained across any intervening material. Maintenance or storage demands are often attributed to be the source of processing costs for structural predictions in sentence comprehension (Gibson, 2000; Wagers & Phillips, 2014; see also; Just & Carpenter, 1992); if activity in IFG and pSTS specifically reflects these maintenance costs, we would not expect increased activation for the two-word phrase condition relative to lists. Importantly for both possibilities, in contrast to several recent studies of syntactic combination using random event-related designs (e.g., Pallier et al., 2011), we used a short block design with cues to the structure before each trial. This design is critical because in most random event-related designs the degree of structure of each item is

uncertain while being processed. If subjects encounter a two-word phrase in this context, they might generate and maintain predictions for a longer and more complex structure before realizing that the current structure only spans two words. By contrast, in our study there is no such uncertainty – therefore we did not expect subjects to predict larger structures than actually present in each stimulus. We also designed our materials in a way that perfectly matched lexical content in the phrase condition with the sentence and list conditions, which is rare among recent work on short phrases but important for ruling out low-level stimulus effects.

In summary, according to the hypothesis that left IFG and pSTS support basic syntactic combination operations, we would expect to see increased activity for both sentences and phrases relative to unstructured lists. If, on the other hand, these regions index non-combinatorial top-down processes that facilitate sentence processing (such as structural prediction), we would expect to see increased activation for the sentence condition only, with no increased activation for phrases relative to word lists. Finally, any increased activity in IFG for our simple sentences compared to lists would add to the evidence that activity for sentences in this region cannot be attributed solely to processing complex or difficult input (e.g., Fedorenko, Duncan, et al., 2012; Pallier et al., 2011; Zaccarella et al., 2015).

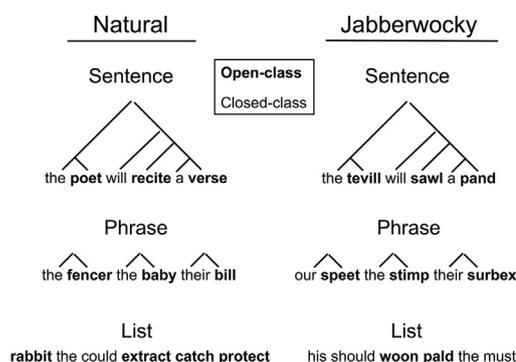
## 2. Materials and methods

### 2.1. Subjects

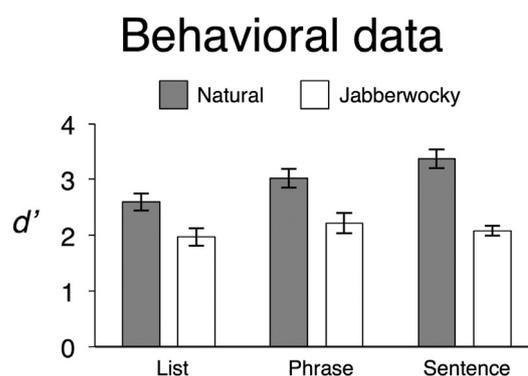
16 subjects were included in the analyses presented here (9 women) (mean age: 24, range: 20–29). One additional subject was excluded for excessive motion during fMRI scanning. Subjects were right-handed, native speakers of English, with normal or corrected-to-normal vision. These subjects participated in both the fMRI experiment and a magnetoencephalography (MEG) experiment with a nearly identical design (we report the MEG data elsewhere). The overlap between the words used for stimuli in the fMRI and MEG sessions was minimal as described below. The order of fMRI and MEG sessions was counterbalanced across subjects; sessions were separated by at least 2 days (mean: 12 days, standard deviation: 9 days). Consent was acquired from each subject before the study began, separately for both the fMRI and MEG sessions, and all procedures were approved by the Institutional Review Board of the University of Maryland.

### 2.2. Stimuli

The current study comprised a core 3 (STRUCTURE)  $\times$  2 (CONTENT) design (Fig. 1). Stimuli consisted of six-word sequences, with one of three different levels of structure: (i) no structure (LIST condition), (ii) intermediate structure, (PHRASE condition), and (iii) full structure (SENTENCE condition). The two levels of CONTENT were: (i) stimuli with real open-class or content words (NATURAL) and (ii) stimuli with pseudo-content words (JABBERWOCKY). Both NATURAL and JABBERWOCKY conditions included real closed-class or functional words.



**Fig. 1 – Schematic of stimulus design. Tree diagrams represent the constituent structure in each condition.**



**Fig. 2 – Mean performance in  $d'$  for each condition for open-class probes. Error bars indicate standard error of the mean with between-subject variability removed (Cousineau, 2005).**

We first generated 432 unique real sentences in three balanced groups of 144 sentences. Each sentence conformed to the same simple, active structure, in which the subject and object always consisted of a determiner (Det)-noun sequence, the sentence always contained an overt modal auxiliary, and the verb was always used transitively: [Det Noun Modal Verb Det Noun], e.g., *the poet will recite a verse*. In order to reduce overt morphological complexity that might induce structure-building operations, the content words (nouns and verbs) in the natural condition bore no overt inflectional morphology (i.e., all singular nouns, no tense/agreement on the verb). Derivational morphology was limited but unavoidable in the NATURAL condition (e.g., “er” suffix on nouns, e.g., *smuggl-er*). However, this derivational morphology was removed from the pseudo-content words in the JABBERWOCKY condition. The set of determiners included articles (*a/an, the*) and possessive pronouns (*his, our, their, my, your*), and the set of modal auxiliaries included *could, should, would, might, must, and will*. In general, we ensured that sentences were semantically coherent and did not violate any selection restrictions, and restricted the maximum length of words to four syllables. We greatly minimized repetition of content words – no content word was repeated more than twice in each group, with no more than four repetitions of any individual word across groups. We grouped together sets of six sentences (a total of 36 words) to

create individual stimulus blocks. These blocks comprised the NATURAL SENTENCE condition.

For the JABBERWOCKY SENTENCE condition, we modified the content words from the whole set of NATURAL sentences using the Wuggy software (Keuleers & Brysbaert, 2010), to create pseudo-content words that were pronounceable and had subsyllabic structure and transitional probabilities that were similar to the real words, e.g., *the tevill will sawl a pand*. We next scrambled the position of these pseudo-content words across the sentences within each 144-sentence group, keeping the positions of the function words fixed. This eliminated superficial similarities to the NATURAL sentences. We then grouped random sets of six of these JABBERWOCKY sentences into individual stimulus blocks.

To create the PHRASE conditions, we pulled out pairs of function and content words from the SENTENCE conditions (e.g., Det-Noun and Modal-Verb), separately for NATURAL and JABBERWOCKY, and created sequences that consisted of three isolated phrases of the same type: [Det-Noun Det-Noun Det-Noun] and [Modal-Verb Modal-Verb Modal-Verb]. Each PHRASE block consisted of four six-word sequences of Det-Noun and two six-word sequences of Modal-Verb, with order of Det-Noun and Modal-Verb sequences randomized within the block.

Finally, we created the LIST conditions, separately for NATURAL and JABBERWOCKY, by randomly permuting the words across all 36 words within the individual SENTENCE blocks, creating six six-word sequences that contained little or no phrasal structure. Any instance of a content word following a function word did create an occasional phrase, but there was drastically less structure in the LIST condition than in the PHRASE condition. If two phrases were consecutive within a sequence (e.g., *might modify the party*), items were manually rearranged.

Our stimuli therefore comprised three groups, each group containing 1/3 of the sentences, phrases, and lists, with the lexical material matched across them (likewise for jabberwocky). We then created 3 different stimulus protocols that combined sentences, phrases and lists from different groups to ensure that there was no lexical overlap between any of the conditions for each protocol. Subjects were presented with different protocols across the fMRI and MEG experiments; thus there was some lexical overlap for each subject (e.g., one subject received blocks of phrases in the MEG experiment that contained the same lexical items as blocks of sentences in the fMRI experiment). Given the scrambling of lexical items across positions throughout each stimulus block and the time in-between scans, we believe any repetition effects were negligible. In addition, this design balanced lexical items across conditions across subjects to control for any specific effects of lexical items. All of our stimuli can be found in [Appendix A](#).

### 2.3. Language probe task

Probe words were presented following 2/6 sequences in each block. These sequences were selected randomly with the stipulation that at least one of the last two sequences had a probe word – this ensured that subjects would not encounter two probes early in the block and then stop paying attention. Probes matched a word within the sequence 50% of the time.

Probes from within the sequence were randomly chosen with equal frequency across different positions in the sequence, and probes not from the sequence were chosen at random from a separate list of open-class items that were never used in the sequences. We did not present closed-class lure items because of the limited set of closed-class items in the experiment. Because closed-class items were frequently repeated, presenting new closed-class items might be too easy, and presenting closed-class items from within our set but not in the targeted sequence might be too difficult. However, we retained closed-class probes from within the sequence to ensure attention to each word. We asked subjects to press a button indicating whether or not they recognized the word from the immediately preceding sequence, and to make their response after the onset of the probe and before the onset of fixation for the next sequence. Subjects made their “yes” responses with one hand and “no” responses with the other, counterbalanced across subjects.

### 2.4. Active baseline condition: counting

We included a non-linguistic active baseline condition. This allowed us to obtain approximate measures of response strength in particular areas by subtracting out the response to an active task that would be less likely to include default mode network activity (Binder et al., 2011). The presentation of the MATH condition was very similar to the language conditions, with the difference being the content of the stimuli and the nature of the probe task. Subjects were presented with a sequence of ‘0’s and ‘#’s, with the ‘#’ indicating that the subject should add 1 to their current total, ‘0’ indicating to not add 1. Following the last item in a MATH trial, subjects were presented with a probe number, (e.g., ‘3’), and they were asked to indicate with a button press whether this matched the total of the preceding sequence.

### 2.5. Procedure

We explained to subjects that they would see blocks of stimuli that would sometimes include all real words, which we described to them as “English”, and sometimes they would include nonsense words that were called “jabberwocky”. We gave them example items, and in the scanner they completed a practice run to become familiar with the stimuli and task before starting the real experiment.

The experiment consisted of 20 blocks from each of the 8 conditions (6 language conditions, MATH, and REST), for a total of 160 stimulus blocks, presented in 10 runs (2 blocks per condition per run). We presented blocks of different conditions in random order except that every run ended with a REST block in order to allow the hemodynamic response to return to baseline. We gave subjects a short break in-between runs. The experiment lasted approximately 1.75 h in total. This was longer than many fMRI experiments, but most subjects were able to complete the entire experiment except for two subjects: one completed 7/10 runs, and another completed 9/10 runs.

We presented the experiment visually using the PsychToolBox Matlab package (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Each trial block began with a fixation cross

presented for 1350 msec, followed by 150 msec of blank screen, and then a condition cue for that block (e.g., “ENGLISH SENTENCE”, “JABBERWOCKY LIST”), which appeared on the screen for 800 msec followed by 150 msec of blank screen. We presented these cues so that subjects would be prepared for that condition and adapt their expectations appropriately. The condition cue was followed by 350 msec of fixation and 150 msec of blank screen, and then the first sequence of the block started. Each item sequence began with a fixation cross that appeared for 650 msec, followed by 150 msec of blank screen. All items (words or numbers) were presented with rapid serial visual presentation, with each item appearing for 350 msec with 150 msec of following blank screen (i.e., 500 msec per item). On probe trials (2/6 sequences per block), the last item of the sequence was followed by 200 msec of fixation, 100 msec of blank screen, and the probe with a question mark (e.g., *should? 4?*) which appeared on the screen for 1700 msec. Each block lasted a total of 30 sec. The REST condition consisted only of the condition cue (“REST”), followed by fixation for 14 sec, and then a numerical countdown of 5 sec (to prepare subjects for the next block), for a total of 22 sec.

## 2.6. Behavioral data analysis

We statistically analyzed performance on open-class items only, as open-class probes contained both signal trials (items from within the stimulus) and noise trials (items not within the stimulus), while closed-class probes were always signal trials (i.e., we never presented a closed-class item that wasn't in the preceding sequence). Thus we could only calculate  $d'$  values that correct for response bias for open class items (we report unanalyzed percent correct on closed-class probes). We performed a  $3 \times 2$  ANOVA for the factors STRUCTURE and CONTENT for open-class probes. We predicted that both structure and content would facilitate performance on the task, in accord with the expectation that chunking due to structure would facilitate recall of individual items, and that real items would be remembered better than pseudo-word items.

## 2.7. fMRI data collection and whole-brain analysis

MR images were obtained in a Siemens TRIO 3T scanner (Siemens Medical Systems) using a 32-channel head coil. We first collected a high-resolution T1-weighted anatomical image in the axial plane (voxel dimension: .45 mm  $\times$  .45 mm  $\times$  .9 mm). We then collected a total of 2370 T2\*-weighted EPI volumes over 10 runs using EPI. Each volume contained 36 oblique slices oriented approximately 20° clockwise relative to the AC-PC axis (TR = 2 sec, TE = 25 msec, flip angle = 90°, in-plane resolution = 3 mm  $\times$  3 mm, slice thickness = 3 mm with .3 mm gap). The first four volumes of each run were collected before stimulus presentation and discarded to control for T1 saturation effects. Slice-timing correction, motion correction, and spatial smoothing were performed using AFNI software (Cox, 1996; <http://afni.nimh.nih.gov/afni>). Motion correction was achieved by using a 6-parameter rigid-body transformation, with each functional volume in each run first aligned to a single volume in that run. Functional volumes were aligned to the anatomical image,

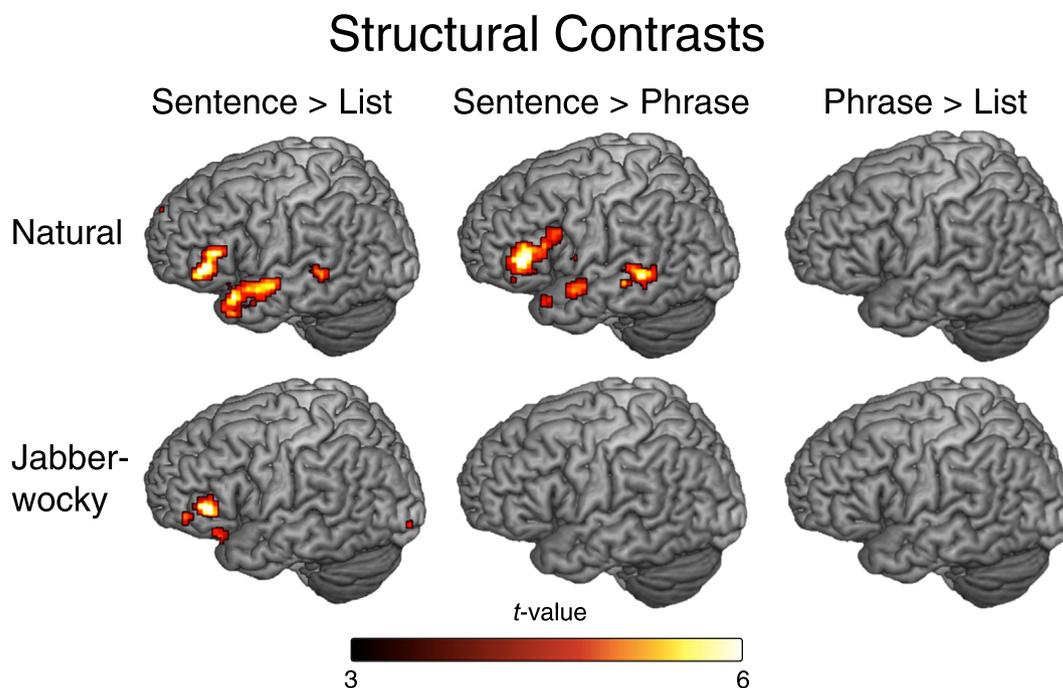
and subsequently aligned to Talairach space (Talairach & Tournoux, 1988, p. 122). Functional images were resampled to 3 mm isotropic voxels, and spatially smoothed using a Gaussian kernel of 6 mm FWHM. The data were high-pass filtered with a cutoff frequency of .0056 Hz at the first-level analysis stage using AFNI's 3dDeconvolve function using the 'polort' parameter with a value of 4.

First-level analyses were performed on each individual's data using AFNI's 3dDeconvolve function. The regression analysis was performed to find parameter estimates that best explained variability in the data. Each predictor variable representing the timecourse of stimulus presentation was entered into a convolution analysis using a canonical hemodynamic response function (AFNI's BLOCK parameter). The following seven regressors of interest were used in the experimental analysis: NATURAL SENTENCE, JABBERWOCKY SENTENCE, NATURAL PHRASE, JABBERWOCKY PHRASE, NATURAL LIST, JABBERWOCKY LIST, and MATH. The six motion parameters were included as regressors of no interest.

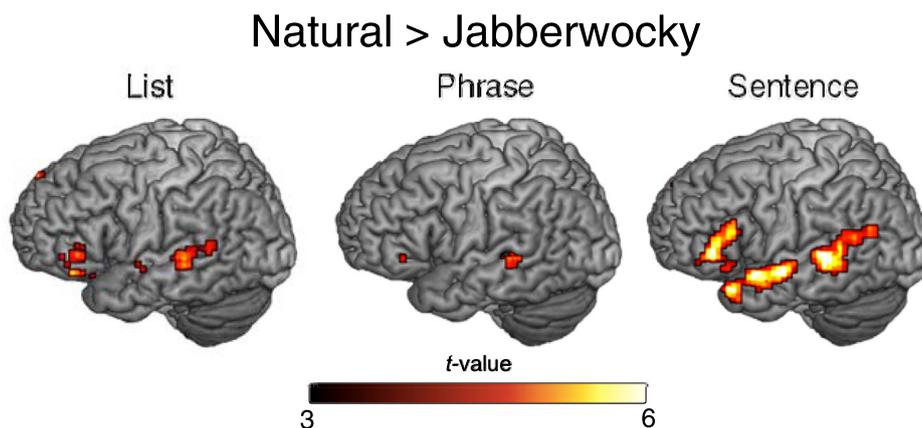
To perform the group-level analysis, we entered the parameter estimates for percent signal change for each condition for each subject into AFNI's 3dANOVA2 function. To correct for multiple comparisons, we used a family-wise error (FWE) cluster size correction using Monte Carlo simulations to estimate the Type 1 error rate for spatial correlation in the data with AFNI's AlphaSim function. We adopted a voxel-wise statistical threshold of  $p < .001$  (one-tailed) and a cluster extent threshold of 31 voxels (837 mm<sup>3</sup>) to keep the FWE rate at  $p < .05$ . For visualization of whole-brain contrasts (Figs. 3 and 4) we used an uncorrected voxel-wise threshold of  $p < .001$  (one-tailed).

## 2.8. ROI analyses

We first delineated regions of interest (ROIs) based on the coordinates of ROIs reported in Pallier et al. (2011). We chose these ROIs because our study most closely followed their design, and their study identified several brain regions whose activity was very sensitive to degree of structure, including the main regions we desired to examine in our study: the IFG and pSTS. There were six ROIs: the anterior superior temporal sulcus (aSTS), the posterior superior temporal sulcus (pSTS), the angular gyrus or temporal–parietal junction (AG/TPJ), the temporal pole (TP), the anterior/inferior portion of the IFG (*pars orbitalis*), and the middle/superior portion of the IFG (*pars triangularis*). We first converted their coordinates from MNI space to Talairach space using AFNI. We defined a sphere around each coordinate with a radius of 10 mm. To ensure that only functionally relevant voxels were included in each ROI (rather than inactive grey matter or white matter), we intersected these spheres with the results of the whole-brain contrast of NATURAL SENTENCES > MATH, thresholded at  $p < .001$  (one-tailed, uncorrected). We assumed that this contrast would activate the language network most robustly, creating the most inclusive ROIs relevant to detecting effects of STRUCTURE while still excluding inactive voxels from the spheres. We then extracted the average percent signal change values for each of the conditions for each of the ROIs. We analyzed effects of STRUCTURE within each ROI, treating individual ROIs as families for purposes of multiple comparisons corrections. We



**Fig. 3** – Group activation maps displayed in Talairach space on a template brain for the pairwise comparisons of **STRUCTURE** (all images display left hemisphere clusters). Activations are displayed at uncorrected  $p < .001$  (one-tailed).



**Fig. 4** – Group activation maps displayed in Talairach space on a template brain for the pairwise comparisons of **CONTENT** (all images display left hemisphere clusters). Activations are displayed at uncorrected  $p < .001$  (one-tailed).

note that we partially defined these ROIs by using data from our statistical analysis within these ROIs, but we justify our choices by (i) the fact that our ROIs were intersections of coordinate-defined regions with our functional contrast, which refines the boundaries of these functionally-independent anatomical areas of interest, and (ii) the fact that our analyses were aimed at determining whether the phrase > list contrast would activate sentence-sensitive regions, and the phrase versus list contrast is orthogonal to our ROI refinement contrast. Since increased activation for sentences relative to unstructured conditions is a somewhat common finding in the literature (e.g., [Bedny, Pascual-Leone, Dodell-Feder, Fedorenko, & Saxe, 2011](#); [Fedorenko, Duncan, et al., 2012](#); [Fedorenko, Nieto-Castanon, & Kanwisher, 2012](#);

[Pallier et al., 2011](#)), our main conclusions stand on the results of the phrase versus list contrast.

One reviewer pointed out that although [Pallier et al. \(2011\)](#) did not obtain significant activation in the *pars opercularis* of Broca's area, previous research has associated this region with syntactic processing ([Friederici, 2016](#)), and therefore it is important to test this region in the present study. We initially followed the approach of [Pattamadilok, Dehaene, and Pallier \(2016\)](#), who took anatomical coordinates from [Amunts et al. \(2004\)](#). When we used these coordinates, created a sphere of 10 mm radius and intersected it with the Natural sentence > MATH contrast, we only obtained a small ROI of 14 voxels that partially overlapped with the *pars triangularis* ROI. We considered this unsuitable for analysis, and so instead used

the cytoarchitectonic maps provided in AFNI for BA44 (Amunts et al., 1999), which strongly overlaps with the *pars opercularis*.

Given that we defined our ROIs using the sentence condition, to ensure that we did not exclude voxels that preferred lists or phrases relative to sentences in our ROIs, we performed exploratory whole-brain analyses at reduced thresholds. NATURAL LIST > NATURAL SENTENCE identified only a ~5 voxels in the *pars opercularis* at a very liberal threshold  $p < .01$  (one-tailed) (these voxels were included in our BA44/*pars opercularis* ROI). NATURAL PHRASE > NATURAL SENTENCE identified no voxels in the vicinity of the relevant regions even at this reduced threshold. JABBERWOCKY LIST > JABBERWOCKY SENTENCE identified no voxels in the vicinity of our regions of interest at  $p < .01$  (one-tailed), nor did the contrast of JABBERWOCKY PHRASE > JABBERWOCKY SENTENCE. These analyses confirmed that the NATURAL SENTENCE > MATH contrast produced the most inclusive ROIs and that we didn't exclude any voxels that preferred phrases or lists relative to sentences.

### 3. Results

#### 3.1. Behavioral data

Performance on the open-class items in the behavioral task for each condition is displayed in Figure 2. A  $3 \times 2$  ANOVA showed a significant main effect of STRUCTURE:  $F(2, 30) = 4.378$ ,  $p = .021$ , a significant main effect of CONTENT:  $F(1,15) = 23.082$ ,  $p < .001$ ; the interaction only trended toward significance:  $F(2,30) = 2.668$ ,  $p = .086$ . We then performed follow-up one-tailed t-tests to determine which levels of structure exhibited significant effects (corrected for multiple comparisons with a family-wise alpha threshold of  $p < .05$ , using a Bonferroni correction with an individual threshold of  $p < .0125$ ): NATURAL PHRASE > NATURAL LIST:  $t(15) = 1.968$ ,  $p = .034$ ; JABBERWOCKY PHRASE > JABBERWOCKY LIST:  $t(15) = 1.042$ ,  $p = .157$ ; NATURAL SENTENCE > NATURAL PHRASE:  $t(15) = 1.522$ ,  $p = .074$ ; JABBERWOCKY SENTENCE > JABBERWOCKY PHRASE:  $t(15) = -.682$ ,  $p = .747$ . None of these effects were significant at the corrected threshold, but there was a trend toward significance of the natural phrase condition relative to natural lists and natural sentences relative to natural phrases. Overall, the ANOVA results confirm our expectations that STRUCTURE and CONTENT impacted task performance, and while the pairwise differences among the conditions were not reliable there appeared to be a progressive effect of STRUCTURE on task performance for NATURAL stimuli.

The average performance on the MATH active baseline task was a  $d'$  of 2.055 (SEM: .274), indicating that subjects attended to this condition. For the closed-class probes, subjects also performed well, indicating that they successfully paid attention to closed-class items throughout the experiment. The average percent correctly identified closed-class probes for each condition are the following: NATURAL LIST: mean 87.4%; NATURAL PHRASE: mean 86.2%; NATURAL SENTENCE: mean 89.0%; JABBERWOCKY LIST: 81.2%; JABBERWOCKY PHRASE: 83.5%; JABBERWOCKY SENTENCE: 79.9%.

#### 3.2. Whole-brain contrasts

The whole-brain contrasts of STRUCTURE are displayed in Fig. 3, and significant clusters are reported in Table 1. The NATURAL

SENTENCE > NATURAL LIST and NATURAL SENTENCE > NATURAL PHRASE contrasts revealed similar activation in several left hemisphere language-related regions: the IFG (straddling the *pars orbitalis* and *pars triangularis*; for NATURAL SENTENCE > NATURAL PHRASE only, the *pars opercularis*), the pSTS (this cluster narrowly avoided significance for the contrast of NATURAL SENTENCE > NATURAL LIST), and the ATL, as well as one small cluster in the right cerebellum. The JABBERWOCKY SENTENCE > JABBERWOCKY LIST contrast only revealed activity in the IFG (*pars triangularis*). The PHRASE > LIST whole-brain contrasts did not reveal any activation for NATURAL OR JABBERWOCKY.

The whole-brain contrasts of CONTENT are displayed in Fig. 4. The NATURAL LIST > JABBERWOCKY LIST contrast revealed a small (non-significant) cluster in the IFG (*pars triangularis*) and a cluster in the pSTS. The NATURAL PHRASE > JABBERWOCKY PHRASE contrast revealed only a small cluster in the pSTS. The NATURAL SENTENCE > JABBERWOCKY SENTENCE contrast revealed the most robust effects, with activation in the IFG (*pars orbitalis* and *pars triangularis*), the ATL (TP and aSTS) and the PTL (pSTS and pMTG), extending posteriorly into the vicinity of the AG/TPJ. This contrast also revealed a cluster in the left hippocampus, the right cerebellum, and a small (non-significant) cluster in the right ATL.

#### 3.3. ROI analyses

The percent signal change values averaged across each ROI and condition are displayed in Fig. 5, and the detailed results of statistical analyses are displayed in Tables 2 (ANOVAs) and 3 (pairwise comparisons). In all regions, except for the *pars opercularis*, there was clearly more activation for NATURAL SENTENCE than any other condition, in accord with our expectations from the literature. Our ANOVAs revealed significant effects of STRUCTURE and CONTENT in every region, and a significant interaction in the TP, aSTS, and AG/TPJ. We followed up the main effect of STRUCTURE by performing pairwise comparisons in order to determine whether the effect of STRUCTURE held at the phrase level, and pairwise comparisons for the sentence condition compared to the phrase condition to ensure that the increased activation for sentences was significant. We did these tests separately for NATURAL and JABBERWOCKY, even though we did not obtain a significant interaction between STRUCTURE and CONTENT in the *pars orbitalis*, *pars triangularis*, *pars opercularis*, and pSTS. We did this to confirm effects separately for NATURAL and JABBERWOCKY as well as to allow transparent comparisons across ROIs (see Table 3).

Critically, there were no significant effects of PHRASE > LIST in any regions, either for NATURAL OR JABBERWOCKY. Every ROI showed a significant effect of NATURAL SENTENCE > NATURAL PHRASE. Nearly every ROI also showed a significant effect of NATURAL SENTENCE > NATURAL LIST; the BA44/*pars opercularis* ROI showed no hint of NATURAL SENTENCE > NATURAL LIST, and this effect just missed significance in AG/TPJ. For JABBERWOCKY SENTENCE > JABBERWOCKY PHRASE, the *pars triangularis*, *pars orbitalis*, *pars opercularis*, and pSTS all showed a trend toward significance; no other ROIs showed this effect. The *pars triangularis* and *pars orbitalis* both showed a significant effect of JABBERWOCKY SENTENCE > JABBERWOCKY LIST; this effect trended toward significance in the pSTS; no other ROIs showed this effect. These results suggest that the main effect of structure observed in the ANOVAs across

**Table 1 – Significant clusters from whole-brain analyses.**

Region	Hemisphere	x	y	z	Cluster size (mm <sup>3</sup> )
<i>Natural Sentence &gt; Natural List</i>					
Anterior temporal lobe	Left	–47	1	–11	2565 mm <sup>3</sup>
Inferior frontal gyrus	Left	–48	29	7	1998 mm <sup>3</sup>
Pars triangularis (posterior/superior peak)	Left	–50	25	14	
Pars orbitalis (anterior/inferior peak)	Left	–44	31	2	
Medial temporal pole	Left	–32	11	–26	1053 mm <sup>3</sup>
Cerebellum	Right	21	–74	–34	972 mm <sup>3</sup>
<i>Natural sentence &gt; Natural Phrase</i>					
Inferior frontal gyrus (pars triangularis, pars opercularis)	Left	–47	27	12	3294 mm <sup>3</sup>
Posterior middle temporal gyrus	Left	–51	–40	1	2025 mm <sup>3</sup>
Cerebellum	Right	22	–66	–34	1053 mm <sup>3</sup>
Amygdala	Left	–25	–4	–10	837 mm <sup>3</sup>
<i>Jabberwocky Sentence &gt; Jabberwocky List</i>					
Inferior frontal gyrus (pars triangularis)	Left	–48	30	8	891 mm <sup>3</sup>
<i>Natural List &gt; Jabberwocky List</i>					
Superior temporal sulcus	Left	–48	–36	4	1296 mm <sup>3</sup>
<i>Natural Phrase &gt; Jabberwocky Phrase</i>					
Middle temporal gyrus	Left	–46	–36	1	972 mm <sup>3</sup>
<i>Natural Sentence &gt; Jabberwocky Sentence</i>					
Superior temporal sulcus	Left	–48	–44	6	6291 mm <sup>3</sup>
Middle temporal gyrus (anterior peak)	Left	–50	–38	2	
Temporal–parietal junction (posterior-medial peak)	Left	–34	–50	8	
Anterior temporal lobe	Left	–48	–44	6	3645 mm <sup>3</sup>
Temporal pole (anterior peak)	Left	–44	17	–18	
Anterior superior temporal sulcus (middle peak)	Left	–47	5	–9	
Anterior superior temporal sulcus (posterior peak)	Left	–50	–10	–7	
Inferior frontal gyrus (pars triangularis)	Left	–49	24	–13	2619 mm <sup>3</sup>
Cerebellum	Right	21	–69	–31	1647 mm <sup>3</sup>
Hippocampus	Left	–28	–12	–9	972 mm <sup>3</sup>

FWE cluster-corrected at  $p < .05$ , individual voxel threshold  $p < .001$  (one-tailed), cluster size threshold 837 mm<sup>3</sup>. All coordinates are center-of-mass (unless noted as local peaks) reported in Talaraich space (Talairach & Tournoux, 1988).

regions was largely driven by increased activation in the *SENTENCE* condition. Although a few *PHRASE > LIST* effects trended toward significance, there was no indication of any such trend in the *pars triangularis* or pSTS. The *NATURAL PHRASE > NATURAL LIST* trended toward significance in the aSTS and the TP, consistent with the results of Pallier et al. (2011), who observed incremental effects of structure in these regions for natural stimuli, but not for jabberwocky. We also saw a trend toward effects for *NATURAL PHRASE > NATURAL LIST* and *JABBERWOCKY PHRASE > JABBERWOCKY LIST* in the *pars orbitalis*, suggesting that this most anterior part of the IFG may prefer phrases to lists. When the analysis was collapsed across both levels of content, the overall pattern did not change, although the effect of *PHRASE > LIST* in the *pars orbitalis* was statistically significant.

#### 4. Discussion

In this study we sought to address an important puzzle in the search for the neural correlates of basic syntactic combination: why do some studies find activation in the left IFG and pSTS for simple sentences or sub-sentence structures relative to unstructured lists or to rest (Bedny et al., 2011; Pallier et al., 2011; Fedorenko, Behr, & Kanwisher, 2011; Zaccarella et al., 2015) while many other studies do not (Friederici et al., 2010; Humphries et al., 2006; Mazoyer et al., 1993; Rogalsky & Hickok, 2009; Stowe et al., 1998)? Our study is novel in

exhibiting both patterns in the same group of subjects and thus it provides important new insights. The hypothesis that we proposed to resolve this quandary is that increased activity in IFG and pSTS for structure reflects *active prediction* mechanisms that are not necessary for parsing simple structures, but may be triggered in certain contexts, particularly when predictions can improve performance on an experimental task. Our results support this hypothesis by the finding that the IFG and pSTS show increased activity for sentences, which likely involved prediction, but not two-word phrases, which likely did not, while both conditions involved increased structure relative to lists.

We observed increased activity throughout the left hemisphere language network for short, simple sentences relative to unstructured word lists, replicating previous studies (Bedny et al., 2011; Fedorenko et al., 2011; Pallier et al., 2011; Zaccarella et al., 2015). Also consistent with these studies, we found increased activity for jabberwocky sentences relative to jabberwocky lists in the IFG and pSTS. However, we did not observe any increased activation for simple two-word natural or jabberwocky phrases relative to lists in these regions (with the possible exception of the most anterior portion of the IFG, the *pars orbitalis*). The absence of effects for two-word phrases is inconsistent with the hypothesis that these regions support basic syntactic combination (Friederici, 2016; Hagoort, 2005; Pallier et al., 2011), but can be naturally explained by a syntactic prediction account. Our results also add to a growing

## ROI percent signal change plots

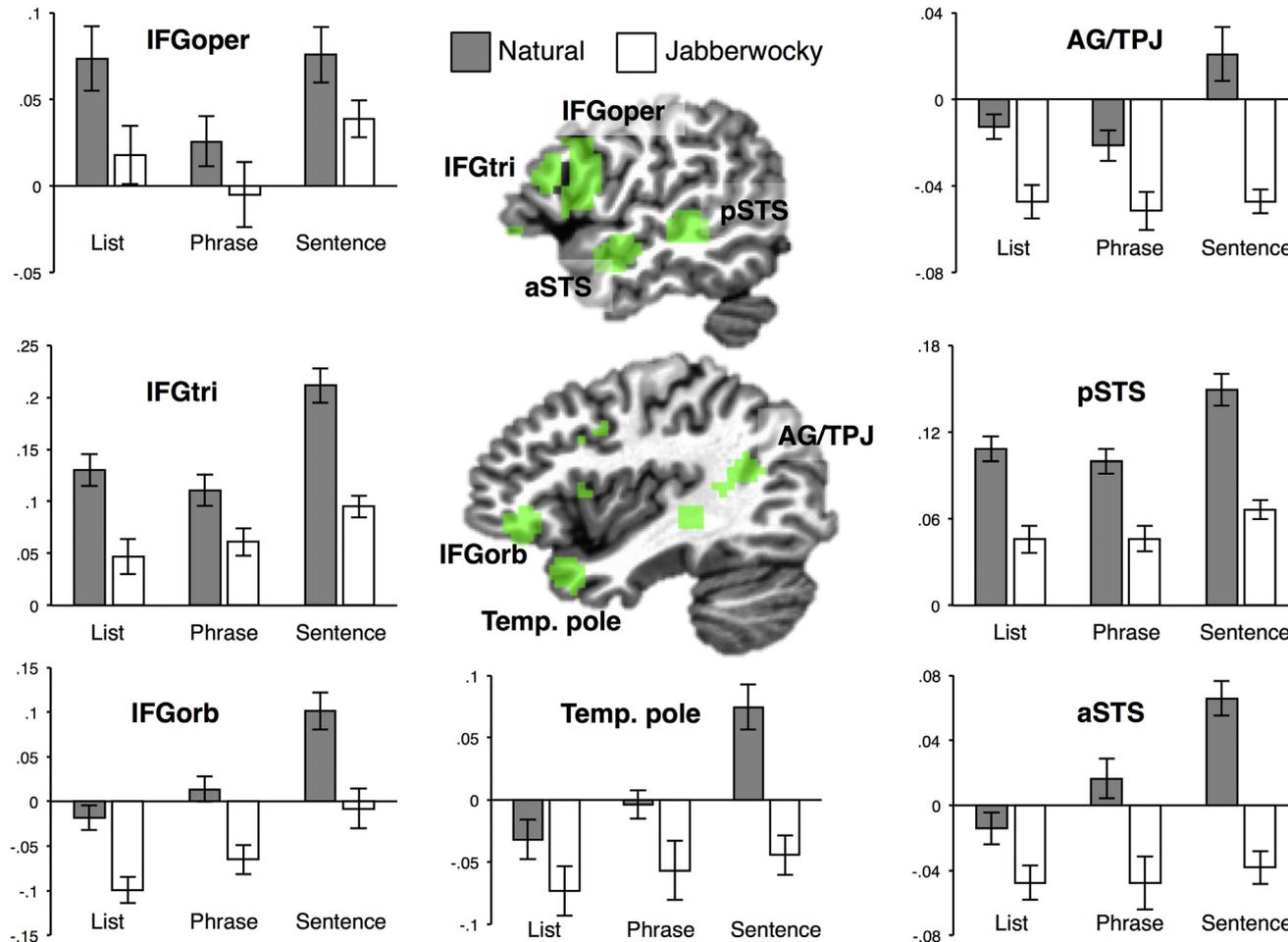


Fig. 5 – Average percent signal change values of each language condition the in left hemisphere ROIs derived from the coordinates reported in [Pallier et al. \(2011\)](#) (with the exception of the *pars opercularis*, derived from the probability map of BA44 created by [Amunts et al., 1999](#)). In the center are the ROIs plotted on individual slices from a template brain in Talairach space, with labels corresponding to each region: *pars opercularis* of the IFG (IFGoper), *pars triangularis* of the IFG (IFGtri), *pars orbitalis* of the IFG (IFGorb), temporal pole (Temp. pole), anterior superior temporal sulcus (aSTS), posterior superior temporal sulcus (pSTS), and angular gyrus (AG/TPJ). Error bars represent the standard error of the mean for each condition with between-subject variability removed (Cousineau, 2005).

**Table 2 – Statistical analyses within each ROI – ANOVA.**

Region	Main effect of STRUCTURE	Main effect of CONTENT	STRUCTURE × CONTENT Interaction
<i>pars orbitalis</i>	<b>F(2,30) = 17.841</b> <b>*p &lt; .001</b>	<b>F(1,15) = 24.566</b> <b>*p ≤ .001</b>	F(2,30) = .372 p = .693
<i>pars triangularis</i>	<b>F(2,30) = 19.444</b> <b>*p &lt; .001</b>	<b>F(1,15) = 21.258</b> <b>*p ≤ .001</b>	<b>F(2,30) = 1.852</b> p = .175
<i>Pars opercularis</i>	<b>F(2,30) = 7.241</b> <b>*p = .003</b>	<b>F(1,15) = 6.372</b> <b>*p = .023</b>	F(2,30) = .181 p = .836
pSTS	<b>F(2,30) = 11.165</b> <b>*p &lt; .001</b>	<b>F(1,15) = 30.201</b> <b>*p &lt; .001</b>	F(2,30) = 1.338 p = .278
Temporal Pole	<b>F(2,30) = 4.910</b> <b>*p = .0143</b>	<b>F(1,15) = 15.701</b> <b>*p = .0013</b>	<b>F(2,30) = 2.873</b> <b>~p = .072</b>
aSTS	<b>F(2,30) = 7.570</b> <b>*p = .0022</b>	<b>F(1,15) = 25.650</b> <b>*p &lt; .001</b>	<b>F(2,30) = 3.563</b> <b>*p = .041</b>
TPJ/AG	<b>F(2,30) = 5.209</b> <b>*p = .0114</b>	<b>F(1,15) = 14.225</b> <b>*p = .0018</b>	<b>F(2,30) = 3.643</b> <b>*p = .038</b>

All t-tests are one-tailed. Significance is corrected for multiple comparisons with a family-wise alpha threshold of  $p < .05$  (each region is defined as a separate family), using a Bonferroni correction with an individual  $p$  threshold of  $p < .0125$ . Tests that are significant or trending towards significance are bolded. \*significant, ~trend toward significance.

**Table 3 – Statistical analyses within each ROI – pairwise comparisons.**

Region	Sentence > List		Sentence > Phrase		Phrase > List	
	Natural	Jabberwocky	Natural	Jabberwocky	Natural	Jabberwocky
<i>pars orbitalis</i>	<b>t(15) = 4.166,</b> <b>*p &lt; .001</b>	<b>t(15) = 2.696,</b> <b>*p = .008</b>	<b>t(15) = 3.740,</b> <b>*p &lt; .001</b>	<b>t(15) = 2.270,</b> <b>~p = .019</b>	<b>t(15) = 1.387,</b> <b>~p = .093</b>	<b>t(15) = 1.396,</b> <b>~p = .092</b>
<i>pars triangularis</i>	<b>t(15) = 3.499,</b> <b>*p = .002</b>	<b>t(15) = 2.758,</b> <b>*p = .007</b>	<b>t(15) = 4.112,</b> <b>*p &lt; .001</b>	<b>t(15) = 1.895,</b> <b>~p = .039</b>	<b>t(15) = -.915,</b> p = .813	<b>t(15) = .645,</b> p = .264
<i>Pars opercularis</i>	<b>t(15) = .076,</b> p = .470	<b>t(15) = 1.099,</b> p = .145	<b>t(15) = 2.634,</b> <b>*p = .009</b>	<b>t(15) = 1.943,</b> <b>~p = .036</b>	<b>t(15) = -1.766,</b> p = .951	<b>t(15) = -.742,</b> p = .765
pSTS	<b>t(15) = 2.687,</b> <b>*p = .008</b>	<b>t(15) = 1.981,</b> <b>~p = .033</b>	<b>t(15) = 4.225,</b> <b>*p &lt; .001</b>	<b>t(15) = 1.875,</b> <b>~p = .040</b>	<b>t(15) = -.662,</b> p = .741	<b>t(15) = .032,</b> p = .487
Temporal Pole	<b>t(15) = 4.117,</b> <b>*p &lt; .001</b>	<b>t(15) = .845,</b> p = .206	<b>t(15) = 3.232,</b> <b>*p = .003</b>	<b>t(15) = .471,</b> p = .322	<b>t(15) = 1.625,</b> <b>~p = .062</b>	<b>t(15) = .454,</b> p = .328
aSTS	<b>t(15) = 5.409,</b> <b>*p &lt; .001</b>	<b>t(15) = .576,</b> p = .287	<b>t(15) = 3.295,</b> <b>*p = .002</b>	<b>t(15) = .474,</b> p = .321	<b>t(15) = 1.735,</b> <b>~p = .052</b>	<b>t(15) = -.012,</b> p = .505
TPJ/AG	<b>t(15) = 2.572,</b> <b>~p = .011</b>	<b>t(15) = .010,</b> p = .496	<b>t(15) = 3.022,</b> <b>*p = .004</b>	<b>t(15) = .541,</b> p = .298	<b>t(15) = -.940,</b> p = .819	<b>t(15) = -.383,</b> p = .646

All t-tests are one-tailed. Significance is corrected for multiple comparisons with a family-wise alpha threshold of  $p < .05$  (each region is defined as a separate family), using a Bonferroni correction with an individual  $p$  threshold of  $p < .0083$ . Tests that are significant or trending towards significance are bolded. \*significant, ~trend toward significance.

body of evidence that the IFG and pSTS play a role in sentence processing that cannot be solely attributed to domain-general mechanisms such as phonological working memory or cognitive control deployed to handle structural complexity, ambiguity, or noncanonicity (Novick et al., 2005; Rogalsky et al., 2008).

**4.1. Prediction in the sentence and phrase conditions**

The argument that our data support the prediction hypothesis relies on two key assumptions: (1) that subjects engaged in top-down structural predictions in the sentence conditions, and (2) that two-word phrases did not elicit effects of prediction, either because subjects did not engage in top-down structural predictions in the phrase conditions or because prediction was less costly in these conditions. We do not have direct evidence of prediction in this experiment, but these assumptions are strongly motivated by the

psycholinguistic literature on sentence processing, which we review below.

According to our hypothesis, structured representations can be built in a strictly bottom-up fashion without predicting structure, but making structural predictions through top-down mechanisms speeds up parsing (and subsequently, semantic interpretation) and improves accuracy of comprehension. Our sentences were short, right-branching (after the initial noun phrase), and contained strictly local dependencies; this obviates the need for selection or maintenance mechanisms that may be invoked for structural complexity or ambiguity. On the other hand, top-down processes can be executed for both complex and simple sentences, making available the use of prediction mechanisms in the sentences in our study. For example, the determiner beginning the first noun phrase in our sentences may trigger a local prediction for the upcoming noun as well as structural nodes and syntactic features (e.g., agreement and tense features) for

the upcoming verb phrase and noun phrase complement. Many such structural predictions are likely cued on the basis of functional elements and therefore would be available in jabbawocky sentences as well as natural sentences (although content words may also contribute to these structural predictions, particularly verbs with subcategorization information). While our experiment did not include a direct index of prediction, we argued in the introduction that subjects were likely to engage predictive mechanisms in the sentence condition, as quickly identifying the structure and content of a sentence would subsequently facilitate word recall, as has been shown in numerous experiments with a task demand very similar to ours (Brener, 1940; Marks & Miller, 1964; Miller & Isard, 1963; Miller et al., 1951).

In contrast, we hypothesized that the two-word phrase condition would not engender predictive effects in IFG/pSTS for at least two reasons:

- (1) If IFG and pSTS specifically index costs for *maintaining* predictions over intervening material, the two-word phrases without intervening material would engender little cost.
- (2) Two-word phrases in our study would be unlikely to elicit structural predictions, because of properties of the experimental presentation and task demands.

According to (1), activity in IFG and pSTS specifically indexes the cost of *maintaining* a prediction across multiple positions. We note here that previous authors have associated processing costs in sentence comprehension to maintenance costs and/or interference costs for intervening material (Gibson, 2000; Lewis, Vasishth, & Van Dyke, 2006; Wagers & Phillips, 2014). Activation in the IFG and pSTS can also be driven by interference between two dependent elements (Glaser, Martin, Van Dyke, Hamilton, & Tan, 2013), suggesting that the difficulty of maintenance might be what really drives activity in IFG/pSTS rather than the generation of predictions themselves. Therefore, the absence of a two-word effect might indicate not that predictions weren't made, but that there was no maintenance cost due to the lack of intervening items.

According to (2), the absence of differences in activity between the two-word phrase condition and the list condition in our experiment indicates that syntactic predictions were not generated in the two-word phrase condition at all. We hypothesized that there is minimal utility in making predictions in this context and with these task demands (the structure is as simple as it gets, and thus easy to accurately process without the benefit of top-down mechanisms). The role of task demands in gating prediction is central for our hypothesis because previous results suggest that predictions are sometimes generated in two-word contexts: both lexical-semantic predictions (Lau, Holcomb, & Kuperberg, 2013) and syntactic predictions (Zaccarella & Friederici, 2015). We argue that these studies used tasks that strongly encouraged prediction. For example, the critical contrast in Zaccarella and Friederici (2015) was between two-word jabbawocky phrases (e.g., *your flirt*) and two-word lists (e.g., *apple flirt*). The task required an immediate judgment as to whether the two words formed a licit phrase or not, and thus predicting the syntactic category could naturally be used as a means to speed responses.

Indeed, Zaccarella and Friederici showed quicker reaction times for phrases relative to lists (supplemental data), and marginally significant increased activation in their phrase condition relative to lists for IFG in BA44, on the border with BA45, consistent with syntactic prediction. In our study, by contrast, behavioral responses were made at the end of a *sequence* of three phrases and a subsequent fixation cross, there was no pressure for subjects to immediately respond, and the task did not directly probe syntactic well-formedness. It is therefore unclear what processing benefits would be gained from prediction in our 2-word condition, whereas in our sentence condition syntactic prediction would facilitate parsing and interpretation which could in turn aid in the memory probe task.

Another important issue given the high regularity of our materials is whether subjects generated non-local predictions in the phrase condition (e.g., at the first determiner, predicted the structure of the subsequent two noun phrases). We think this is unlikely because while the structure of non-local phrases was in principle predictable, the notions of *incrementality* and *connectedness* that are prevalent in the psycholinguistics parsing literature (Gibson, 1998; Stabler, 1994; Sturt & Lombardo, 2005) indicate that people deploy cognitive processes within a single, connected structure rather than generate expectations for unconnected upcoming structure. Maintenance of a single structure allows efficient maintenance of information, while generating and maintaining non-local predictions increases the number of items in working memory. In essence, there would be a substantial memory cost for generating these unconnected, non-local predictions, (at the first noun phrase, two full additional noun phrases) and little to gain by generating them. However, we cannot directly rule out this possibility, and the general issue of whether subjects always generate predictions for stimuli that are in principle predictable is an open one.

#### 4.2. Top-down structural predictions and the broader literature

Much of the variation across experiments in terms of IFG and pSTS activation can be explained by the extent to which differences in stimuli, presentation, and task encouraged subjects to invoke active prediction mechanisms. According to our hypothesis, bottom-up structure-building mechanisms are adequate for simple sentence processing, and solely using these mechanisms is less costly than engaging these top-down processing mechanisms. However, not using active prediction may increase the likelihood of error or “falling behind” for even simple sentence input, and may be inadequate altogether for processing certain complex constructions. Thus, subjects will engage active processing for complex constructions or when task demands put a premium on rapid, accurate parsing of structure; conversely, subjects should be less likely to engage these mechanisms when task demands are light and structures are simple.

This hypothesis accurately distinguishes which neuroimaging studies report increased activity for structured stimuli compared to controls and those that do not. For instance, Mazoyer et al. (1993) and Stowe et al. (1998) were PET studies with long blocks of the same stimulus type (~1.5 min per

block) and no task; consistent with the active prediction hypothesis, they failed to find basic effects of structure in IFG.<sup>1</sup> A clear example of task-dependence of activation in the IFG is the experiment reported by Rogalsky and Hickok (2009). They found little activity for sentences or lists in the IFG and pSTS in a passive listening task; however, when subjects were required to attend to either the syntax or semantics of sentences, the response to sentences increased in both regions, particularly strikingly in the IFG (see their Fig. 2). Some previous failures to observe structural effects in the IFG and/or pSTS may reflect task demands that trigger active prediction not only in the structured condition but also in the *unstructured* condition. A compelling example of this is the study by Humphries et al. (2006), Humphries, Binder, Medler, and Lieberthal (2007), which presented semantically congruent sentences (e.g., *the man on vacation lost a bag and a wallet*), semantically anomalous sentences (e.g., *the freeway on a pie watched a house and a window*), and jaberwocky sentences, as well as scrambled list versions of each (e.g., *on vacation lost then a and bag wallet man then a*). The subjects' task was to judge the meaningfulness of every stimulus along a scale from 1 to 4 during scanning. While Humphries et al. found no differences between sentences and word lists in the IFG or pSTS, in this study both sentences and lists robustly activated these regions above baseline. This result can be explained by subjects using top-down processing in both sentence and lists conditions in order to evaluate meaningfulness, as imposing structure in a top-down fashion on the semantically congruent list condition could make it possible to generate a coherent interpretation. Indeed, subjects rated semantically congruent lists as more meaningful than semantically anomalous lists even though both had no structure.

Within the studies that have observed effects of simple sentence structure in the IFG and pSTS, the tasks and designs used are likely to have encouraged top-down structure-building. As in the current study, Bedny et al. (2011) and Fedorenko, Nieto-Castanon, et al. (2012) asked subjects to indicate whether a given probe word had been presented in the preceding stimulus (for both sentences and lists). Rapid and accurate structure building in the sentence conditions of these experiments would then enhance sentence comprehension and facilitate performance on the memory probe task, and as such top-down processing is likely to have been used. While Pallier et al. (2011) used a somewhat less demanding overt task (detecting an occasional probe sentence) they also used a randomized, event-related design with stimuli ranging from two-word phrases to full twelve-word sentences. This meant that on every trial, in order to achieve even a somewhat accurate structured representation of the string, subjects had to resolve uncertainty about the structure of each string. We suggest that this design often led to structural predictions beyond the range of their short structured stimuli. Therefore in this study it appears to be the structure of the stimuli that encouraged a top-down processing strategy, where evaluating

the fit between an incoming word and the current structural prediction would allow subjects to more rapidly and accurately detect constituent boundaries. This explains why Pallier et al. observed increased IFG/pSTS in the two-word condition in contrast to the current study, where our use of a block design removed any uncertainty about whether the initial two-word phrase would belong to a larger constituent.

#### 4.3. Top-down processing beyond simple structures

As discussed above, structural effects in IFG and pSTS for simple phrases and sentences are important because they cannot be straightforwardly accounted for by working memory, cognitive control or reanalysis operations required for long-distance dependencies or noncanonical structures. As we reviewed in the introduction, there are some studies that explicitly support our top-down prediction hypothesis of these regions' function (Bonhage et al., 2015; Jakuszeit et al., 2013; Matchin et al., 2014; Santi & Grodzinsky, 2007, 2012). Other studies also support the important component of our claim that active processing is under strategic control. Waters, Caplan, Alpert, and Stanczak (2003) in a PET study presented subjects with subject-relative and object-relative sentences (a classic structural comparison that robustly activates IFG; see Meyer & Friederici, 2016 for a review). They split subjects by processing speed – those that responded more quickly and those that responded more slowly; both groups performed near ceiling at comprehension. Their results showed increased activation in the IFG (bilaterally) in the faster subjects for the object-relative > subject-relative contrast, and no such IFG effects for the slower subjects. These data support the notion that the faster subjects used top-down processing to facilitate performance for the complexity contrast, while slower subjects did not. A recent fMRI study by Pattamadilok et al. (2016) presented sentences that differed on syntactic complexity, followed by comprehension probes that required subjects to syntactically parse the sentences to give correct answers. They found complexity effects in the IFG and pSTS at the probe sentences but not during sentence processing itself. These results indicate that subjects did not engage in top-down processing until forced by the task, and that increased activity in these regions reflects the deployment of strategic top-down processing.

One remaining question is the functional distinction between the IFG and pSTS. While answering this question goes beyond the limited goals of this study, we note that prediction mechanisms may increase demands on lexical processing, such as sustained activation of lexical items during prediction (see Snijders et al., 2009 for a similar proposal). Active processing may be accompanied by increased activity associated with lexical access. Thus we tentatively suggest that the IFG underlies the generation and/or maintenance of predictions, and the pSTS underlies the representation of the words or structural nodes that are predicted.

#### 4.4. Semantic processing as an alternative explanation for our results

Our results argue against the hypothesis that IFG and pSTS support basic structure building, and we have suggested instead that the increased activity observed in these regions

<sup>1</sup> We note that Mazoyer et al. (1993) found activity in IFG for full stories; we assume that the rich semantic content encouraged the use of active processing. Crucially, though, this study did not find activity in IFG for jaberwocky sentences or semantically anomalous sentences.

for sentences versus phrases and lists reflects syntactic prediction. However, another alternative explanation for the effects of structure observed in IFG and pSTS is semantic processing. Our sentences differed from our phrases and lists in containing event structure/argument structure, and it is reasonable to think that semantic representations concerning ‘who did what to whom’ could be constructed even in the absence of lexical semantics in the jabberwocky materials. This is an important possibility that could be explored in future work by examining the processing of two-word phrases containing event structure (e.g., John left). However we note that event structure processing does not explain the results of Bonhage et al. (2015), where the increased activity was specifically associated with the prediction of a syntactic category, and effects were observed for jabberwocky sentences but not for natural sentences. Additionally, other studies have found activity in the IFG (specifically, BA45/pars triangularis) for very simple phrases (Sahin, Pinker, Cash, Schomer, & Halgren, 2009; Tyler, Stamatakis, Post, Randall, & Marslen-Wilson, 2005; Zaccarella et al., 2015) that doesn't appear attributable to combinatorial semantics, particularly at the level of event structure; conversely, different regions (particularly the TPJ/AG) have been implicated in event structure processing (e.g., Boylan, Trueswell, & Thompson-Schill, 2015; Schwartz et al., 2011; Thothathiri et al., 2012).

The fact that natural stimuli elicited more activity than jabberwocky stimuli in IFG and pSTS at all levels of structure (list, phrase, and sentence), consistent with several previous studies (e.g., Bedny et al., 2011; Fedorenko, Nieto-Castanon, et al., 2012; Humphries et al., 2006), might also seem to support the idea that these regions are involved in semantic processing. However, note that this main effect of content is not explained by an event structure account like the one above, as it is an effect of lexical semantic content that is observed in unstructured word lists as well as sentences. Furthermore, other studies show a different pattern: Pallier et al. (2011) found no apparent differences between natural and jabberwocky materials in IFG and pSTS (although they did not perform a direct statistical contrast, the observed effects had approximately the same effect sizes for both conditions), and Bonhage et al. (2015) found more activity for jabberwocky sentences than natural sentences. This overall pattern of results might be accommodated by a prediction account if one assumes that degree of structural prediction is graded, rather than all-or-none as we assumed for convenience above. The assumption would be that some prediction occurs automatically even in unstructured contexts on the basis of single-word syntax, and that in natural materials both the content words and the function words can drive these automatic predictions, while in the jabberwocky materials this is primarily limited to function words. This could explain the main effect of natural > jabberwocky observed in the current study, and variation in the impact of experimental context and task parameters on strategic prediction could explain the variability across studies.

#### 4.5. Effects in the ATL, AG, and pars orbitalis: semantic processing

We found effects of structure in the ATL (i.e., the TP and aSTS) and AG for the natural sentence condition only. In these

regions there were clearly no effects of structure in the jabberwocky condition, and there was increased activity at all structural levels for the effect of lexical content. These data largely replicate Pallier et al. (2011)'s results showing structural sensitivity for natural but not jabberwocky stimuli in ATL and AG, and are consistent with their conclusion that these regions are more likely to be involved in semantic or conceptual processing than syntactic processing. In addition, these results are consistent with a series of fMRI and MEG studies implicating the ATL (and somewhat less frequently the AG) in basic semantic composition (Baron & Osherson, 2011; Bemis & Pykkänen, 2011, 2012; Del Prado & Pykkänen, 2014; Vandenberghe, Nobre, & Price, 2002; Westerlund, Kastner, Al Kaabi, & Pykkänen, 2015).

In our experiment, we observed one weak but suggestive difference between the profile in ATL and AG/TPJ: there was clearly no difference between natural lists and two-word phrases in the AG/TPJ, while there appeared to be a three-level parametric effect of structure in the ATL: the contrast of Natural phrases > Natural lists neared but did not reach significance in both the aSTS and TP, with a clear numerical increase for natural phrases compared to natural lists. This matches the results of Brennan et al. (2012), who found a tight correlation between the closure of phrases during sentence processing and activity in the ATL. This pattern is also consistent with Pallier et al.'s (2011) results, in which the ATL showed a parametric effect of increasing structure and the AG/TPJ showed increased activity only for the full-sentence and near-sentence conditions. In contrast, the AG/TPJ has been argued to support full sentence or discourse level processing (Binder, Desai, Graves, & Conant, 2009; Seghier, 2013), in accord with studies implicating this region in thematic or argument structure processing (Lewis, Poeppel, & Murphy, 2015; Thompson et al., 2007; Thothathiri et al., 2012).

While the pSTS and the *pars triangularis* subregion of the IFG showed no indication of increased activation for phrases compared to lists, there was a non-significant trend in this direction in the more anterior portion of the IFG, the *pars orbitalis*, for both natural and jabberwocky conditions. This region is less traditionally associated with language and syntactic processing (the classic definition of Broca's area does not include the *pars orbitalis*; Amunts et al., 1999). However, there is some evidence that damage to this region results in sentence comprehension impairments (Baldo & Dronkers, 2007; Dronkers et al., 2004), and some neuroimaging studies have highlighted this region's participation in the semantic component of sentence processing (Dapretto and Bookheimer, 1999; Hagoort & Indefrey, 2014). In the current study we saw the same trend in both the jabberwocky condition and the natural condition, and therefore if a semantic computation were driving these effects it would have to be fairly abstract. Overall, because effects in this region were weak, we leave the question of the particular computations supported by *pars orbitalis* to future research.

#### 4.6. The *pars opercularis* – verbal working memory

Our BA44/*pars opercularis* ROI exhibited the main effects of CONTENT and STRUCTURE that was seen in all other regions. However, it was notably distinct from all other ROIs in yielding

increased activation for the list condition relative to phrases; for natural lists, the activation was as strong as for natural sentences. This pattern is inconsistent with what would be expected under a syntactic combination account (Friederici, 2016) – rather, increased activation for sentences relative to phrases, and phrases relative to lists, would be expected. On the other hand, this pattern of activity is somewhat consistent with the expectations of a brain area involved in verbal working memory, as previous authors have suggested (Rogalsky et al., 2008; reviewed in detail in Rogalsky & Hickok, 2011). First, the pars opercularis is implicated as part of a speech production circuit (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009), and articulatory rehearsal is thought to be the operative mechanism for verbal working memory (Baddeley, 2003). Sentence comprehension plausibly relies on verbal working memory because concurrent articulatory rehearsal inhibits the comprehension of both written and auditory sentences (Rogalsky et al., 2008; Waters, Caplan, & Hildebrandt, 1987). The lack of structure and content in the list condition also would seem to require substantial verbal working memory resources, as each word would have to be rehearsed individually. However, the phrase condition did not involve structural hierarchy beyond local two-word phrases, not requiring verbal working memory resources to build, and the burden on verbal working memory is much less for phrases compared to lists because only three phrases needed to be rehearsed rather than six isolated words. Altogether, the verbal working memory hypothesis can explain the activation in BA44/pars opercularis seen in our study, while the syntactic combination hypothesis does not.

#### 4.7. Basic syntactic combination

While our data provide evidence for a role for the IFG and pSTS in predictive top-down syntactic processing, we cannot draw firm conclusions about the neural basis for basic or bottom-up syntactic structure building. This is because the phrase > list contrast did not yield significant activity except trending toward significance in the ATL (aSTS and TP) and the pars orbitalis, and these are regions that previous research suggests are associated with semantic processing and not basic syntax. It appears that our study, in addition to the literature that we have reviewed, has failed to offer promising candidates for cortical areas that underlie basic syntactic combination. This failure deserves attention and explanation.

One possibility is that the BOLD response is simply not the right method for identifying the brain basis for syntactic operations. This may be because the BOLD signal is not sensitive to the neural activity that builds basic syntactic representations or because the cortical area is not the right level or neuroanatomical granularity for syntactic mechanisms. This is a general and serious problem with aligning grammatical theory and neuroscience (Embick & Poeppel, 2015; Poeppel & Embick, 2005) – at what level of granularity should we be looking for the computational and representational implementation of language? It might not in fact be the cortical area as is assumed in most neuroimaging and neuropsychology experiments. Future research should develop and test hypotheses of basic syntactic combination beyond the standard search for a “syntax area”.

## 5. Conclusions

In the present study, we found increased activity for simple, short sentences in the left IFG and pSTS relative to unstructured word lists. This adds to recent evidence that activations in these regions during sentence processing cannot be solely attributed to mechanisms invoked to handle complex or non-canonical sentences. However, our results speak against hypotheses that this region supports basic syntactic combination, as we did not observe a simple effect of structure for a two-word phrase condition. Given the substantial psycholinguistic literature on predictive mechanisms in sentence processing, we hypothesize that increased activation in the IFG and pSTS during the processing of simple structures reflects structural predictions. This hypothesis correctly accounts for the variability in the literature regarding when the IFG and pSTS show basic structural effects and when they do not. Our hypothesis also predicts that these basic effects will be seen in experiments that encourage structural predictions; this hypothesis can be tested in future studies of simple structure processing.

Here is a summary of the claims with respect to sentence processing we made for all the brain areas tested in our ROI analyses in the context of the neuroimaging and neuropsychological literature on these brain regions: BA44/pars opercularis, verbal working memory; pars triangularis, top-down syntactic prediction; pars orbitalis, top-down syntactic prediction and/or semantic processing (unspecified); AG/TPJ, event/propositional semantics; pSTS, top-down syntactic prediction/lexical access; aSTS, basic semantic combination; TP, basic semantic combination.

## Acknowledgments

This research was supported by internal startup funds awarded to Ellen Lau at the University of Maryland. We would like to thank Norbert Hornstein for discussion, and the attendees of the 2015 Society for the Neurobiology of Language Meeting in Chicago for their feedback on this work. We would also like to thank two anonymous reviewers for their thorough comments and suggestions.

## Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.cortex.2016.12.010>.

## REFERENCES

- Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H., & Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. *Journal of Comparative Neurology*, 412(2), 319–341.
- Amunts, K., Weiss, P. H., Mohlberg, H., Pieperhoff, P., Eickhoff, S., Gurd, J. M., et al. (2004). Analysis of neural mechanisms underlying verbal fluency in cytoarchitecturally defined

- stereotaxic space—the roles of Brodmann areas 44 and 45. *NeuroImage*, 22(1), 42–56.
- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829–839.
- Baldo, J. V., & Dronkers, N. F. (2007). Neural correlates of arithmetic and language comprehension: A common substrate? *Neuropsychologia*, 45(2), 229–235.
- Baron, S. G., & Osherson, D. (2011). Evidence for conceptual combination in the left anterior temporal lobe. *NeuroImage*, 55(4), 1847–1852.
- Bedny, M., Pascual-Leone, A., Dodell-Feder, D., Fedorenko, E., & Saxe, R. (2011). Language processing in the occipital cortex of congenitally blind adults. *Proceedings of the National Academy of Sciences*, 108(11), 4429–4434.
- Bemis, D. K., & Pykkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *The Journal of Neuroscience*, 31(8), 2801–2814.
- Bemis, D. K., & Pykkänen, L. (2012). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, 23, 1859–1873.
- 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.
- Binder, J. R., Gross, W. L., Allendorfer, J. B., Bonilha, L., Chapin, J., Edwards, J. C., et al. (2011). Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study. *NeuroImage*, 54(2), 1465–1475.
- Bonhage, C. E., Mueller, J. L., Friederici, A. D., & Fiebach, C. J. (2015). Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex*, 68, 33–47.
- Boylan, C., Trueswell, J. C., & Thompson-Schill, S. L. (2015). Compositionality and the angular gyrus: A multi-voxel similarity analysis of the semantic composition of nouns and verbs. *Neuropsychologia*, 78, 130–141.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology*, 26(5), 467.
- Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pykkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, 120(2), 163–173.
- Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, 3(4), 572–582.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding* (Vol. 6). Cambridge MA: MIT Press.
- Chomsky, N. (1995). *The minimalist program* (Vol. 28). Cambridge, MA: MIT press.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3), 162–173.
- Crain, S., & Fodor, J. D. (1985). How can grammars help parsers. *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*, 94–128.
- Dapretto, M., & Bookheimer, S. Y. (1999). Form and content: Dissociating syntax and semantics in sentence comprehension. *Neuron*, 24(2), 427–432.
- Del Prado, P., & Pykkänen, L. (2014). MEG evidence for conceptual combination but not numeral quantification in the left anterior temporal lobe during language production. *Frontiers in Psychology*, 5, 524.
- Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, 92(1), 145–177.
- Embick, D., Marantz, A., Miyashita, Y., O'Neil, W., & Sakai, K. L. (2000). A syntactic specialization for Broca's area. *Proceedings of the National Academy of Sciences USA*, 97(11), 6150–6154.
- Embick, D., & Poeppel, D. (2015). Towards a computational (ist) neurobiology of language: Correlational, integrated and explanatory neurolinguistics. *Language, Cognition and Neuroscience*, 30(4), 357–366.
- Everaert, M. B., Huybregts, M. A., Chomsky, N., Berwick, R. C., & Bolhuis, J. J. (2015). Structures, not Strings: Linguistics as part of the cognitive sciences. *Trends in Cognitive Sciences*, 19(12), 729–743.
- Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, 108(39), 16428–16433.
- Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, 22(21), 2059–2062.
- Fedorenko, E., Nieto-Castanon, A., & Kanwisher, N. (2012). Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia*, 50(4), 499–513.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4), 291–325.
- Fridriksson, J., Fillmore, P., Guo, D., & Rorden, C. (2015). Chronic Broca's aphasia is caused by damage to Broca's and Wernicke's areas. *Cerebral Cortex*, 25(12), 4689–4696.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78–84.
- Friederici, A. D. (2016). The neuroanatomical pathway model of Language: Syntactic and semantic networks. In G. Hickok, & S. A. Small (Eds.), *The Neurobiology of language*. Elsevier.
- Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. D., & von Cramon, D. Y. (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex*, 16(12), 1709–1717.
- Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping*, 31(3), 448–457.
- Friederici, A. D., Makuuchi, M., & Bahlmann, J. (2009). The role of the posterior superior temporal cortex in sentence comprehension. *NeuroReport*, 20(6), 563–568.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1–76.
- Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, Language, Brain*, 95–126.
- Glaser, Y. G., Martin, R. C., Van Dyke, J. A., Hamilton, A. C., & Tan, Y. (2013). Neural basis of semantic and syntactic interference in sentence comprehension. *Brain and Language*, 126(3), 314–326.
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9(9), 416–423.
- Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience*, 37, 347–362.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.
- Humphries, C., Binder, J. R., Medler, D., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, 18(4), 665–679.
- Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: An fMRI study. *NeuroImage*, 36(3), 924–932.
- 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.
- Jackendoff, R. (2003). Précis of foundations of language: Brain, meaning, grammar, evolution. *Behavioral and Brain Sciences*, 26(06), 651–665.
- Jakuszeit, M., Kotz, S. A., & Hasting, A. S. (2013). Generating predictions: Lesion evidence on the role of left inferior frontal cortex in rapid syntactic analysis. *Cortex*, 49(10), 2861–2874.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122.
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3), 627–633.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, 36(14), 1.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29, 627–645.
- Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, 25(3), 484–502.
- Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, 98(1), 74–88.
- Levy, R. P., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, 68(2), 199–222.
- Lewis, G. A., Poeppel, D., & Murphy, G. L. (2015). The neural bases of taxonomic and thematic conceptual relations: An MEG study. *Neuropsychologia*, 68, 176–189.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454.
- Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, 13(3), 361–392.
- Magnusdottir, S., Fillmore, P., Den Ouden, D. B., Hjaltason, H., Rorden, C., Kjartansson, O., et al. (2013). Damage to left anterior temporal cortex predicts impairment of complex syntactic processing: A lesion-symptom mapping study. *Human Brain Mapping*, 34(10), 2715–2723.
- Marks, L. E., & Miller, G. A. (1964). The role of semantic and syntactic constraints in the memorization of English sentences. *Journal of Verbal Learning and Verbal Behavior*, 3(1), 1–5.
- Matchin, W., Sprouse, J., & Hickok, G. (2014). A structural distance effect for backward anaphora in Broca's area: An fMRI study. *Brain and Language*, 138, 1–11.
- 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.
- Meyer, L., & Friederici, A. D. (2016). Neural systems underlying the processing of complex sentences. In G. Hickok, & S. A. Small (Eds.), *The Neurobiology of language*. Elsevier.
- Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, 41(5), 329.
- Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 2(3), 217–228.
- Mohr, J. P., Pessin, M. S., Finkelstein, S., Funkenstein, H. H., Duncan, G. W., & Davis, K. R. (1978). Broca aphasia Pathologic and clinical. *Neurology*, 28(4), 311–311.
- Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3(2), 151–165.
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 5(3), 263–281.
- Omaki, A., Lau, E. F., White, I. D., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology*, 6.
- Pallier, C., Devauchelle, A. D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, 108(6), 2522–2527.
- Pattamadilok, C., Dehaene, S., & Pallier, C. (2016). A role for left inferior frontal and posterior superior temporal cortex in extracting a syntactic tree from a sentence. *Cortex*, 75, 44–55.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Phillips, C. (1996). *Order and structure*. Doctoral dissertation. Massachusetts Institute of Technology.
- Poeppel, D., & Embick, D. (2005). Defining the relation between linguistics and neuroscience. *Twenty-first Century Psycholinguistics: Four Cornerstones*, 103–118.
- Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. University of Chicago Press.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.
- Rogalsky, C., Almeida, D., Sprouse, J., & Hickok, G. (2015). Sentence processing selectivity in Broca's area: Evident for structure but not syntactic movement. *Language, Cognition and Neuroscience*, 30(10), 1326–1338.
- Rogalsky, C., & Hickok, G. (2011). The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience*, 23(7), 1664–1680.
- 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.
- Rogalsky, C., Matchin, W., & Hickok, G. (2008). Broca's area, sentence comprehension, and working memory: An fMRI study. *Frontiers in Human Neuroscience*, 2.
- Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science*, 326(5951), 445–449.
- Santi, A., & Grodzinsky, Y. (2007). Working memory and syntax interact in Broca's area. *NeuroImage*, 37(1), 8–17.
- Santi, A., & Grodzinsky, Y. (2012). Broca's area and sentence comprehension: A relationship parasitic on dependency, displacement or predictability? *Neuropsychologia*, 50(5), 821–832.
- Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O. K., Dell, G. S., et al. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proceedings of the National Academy of Sciences*, 108(20), 8520–8524.
- Seghier, M. L. (2013). The angular gyrus multiple functions and multiple subdivisions. *The Neuroscientist*, 19(1), 43–61.
- Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex*, 19(7), 1493–1503.
- Stabler, E. P. (1994). The finite connectivity of linguistic structure. *Perspectives on Sentence Processing*, 303–336.

- Staub, A., & Clifton, C., Jr. (2006). Syntactic prediction in language comprehension: Evidence from either... or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 425.
- Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1(3), 227–245.
- 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.
- Sturt, P., & Lombardo, V. (2005). Processing coordinated structures: Incrementality and connectedness. *Cognitive Science*, 29(2), 291–305.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York, NY: Thieme Medical Publishers.
- Thompson, C. K., Bonakdarpour, B., Fix, S. C., Blumenfeld, H. K., Parrish, T. B., Gitelman, D. R., et al. (2007). Neural correlates of verb argument structure processing. *Journal of Cognitive Neuroscience*, 19(11), 1753–1767.
- Thompson, C. K., & Mack, J. E. (2014). Grammatical impairments in PPA. *Aphasiology*, 28(8–9), 1018–1037.
- Thothathiri, M., Kimberg, D. Y., & Schwartz, M. F. (2012). The neural basis of reversible sentence comprehension: Evidence from voxel-based lesion symptom mapping in aphasia. *Journal of Cognitive Neuroscience*, 24(1), 212–222.
- Tyler, L. K., Randall, B., & Stamatakis, E. A. (2008). Cortical differentiation for nouns and verbs depends on grammatical markers. *Journal of Cognitive Neuroscience*, 20(8), 1381–1389.
- Tyler, L. K., Stamatakis, E. A., Post, B., Randall, B., & Marslen-Wilson, W. (2005). Temporal and frontal systems in speech comprehension: An fMRI study of past tense processing. *Neuropsychologia*, 43(13), 1963–1974.
- Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*, 14(4), 550–560.
- Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology*, 67, 1274–1304.
- Waters, G., Caplan, D., Alpert, N., & Stanczak, L. (2003). Individual differences in rCBF correlates of syntactic processing in sentence comprehension: Effects of working memory and speed of processing. *NeuroImage*, 19(1), 101–112.
- Waters, G., Caplan, D., & Hildebrandt, N. (1987). Working memory and written sentence comprehension. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*. Erlbaum.
- Westerlund, M., Kastner, I., Al Kaabi, M., & Pylkkänen, L. (2015). The LATL as locus of composition: MEG evidence from English and Arabic. *Brain and Language*, 141, 124–134.
- Wilson, S. M., Dronkers, N. F., Ogar, J. M., Jang, J., Growdon, M. E., Agosta, F., et al. (2010). Neural correlates of syntactic processing in the nonfluent variant of primary progressive aphasia. *The Journal of Neuroscience*, 30(50), 16845–16854.
- Wilson, S. M., Galantucci, S., Tartaglia, M. C., & Gorno-Tempini, M. L. (2012). The neural basis of syntactic deficits in primary progressive aphasia. *Brain and Language*, 122(3), 190–198.
- Wilson, S. M., & Saygin, A. P. (2004). Grammaticality judgment in aphasia: Deficits are not specific to syntactic structures, aphasic syndromes, or lesion sites. *Journal of Cognitive Neuroscience*, 16(2), 238–252.
- Wulfeck, B., & Bates, E. (1991). Differential sensitivity to errors of agreement and word order in Broca's aphasia. *Journal of Cognitive Neuroscience*, 3(3), 258–272.
- Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes*, 28, 272–302.
- Zaccarella, E., & Friederici, A. D. (2015). Merge in the human brain: A sub-region based functional investigation in the left pars opercularis. *Frontiers in Psychology*, 6.
- Zaccarella, E., Meyer, L., Makuuchi, M., & Friederici, A. D. (2015). Building by syntax: The neural basis of minimal linguistic structures. *Cerebral Cortex*. bhv234.