Received Date: 20-Nov-2015

Revised Date: 27-Jul-2016

Accepted Date: 12-Aug-2016

Article Type: Brief Report

MEG evidence for incremental sentence composition in the anterior temporal lobe

Jonathan R. Brennan[1] and Liina Pylkkänen[2,3,4]

1 Department of Linguistics, University of Michigan

2 Department of Linguistics, New York University

3 Department of Psychology, New York University

4 NYUAD Institute, New York University Abu Dhabi

Keywords: Syntax, Semantics, Language Understanding, Magnetoencephalography

Category: Brief Article

Abstract Word Length: 123

Manuscript Length: 4131

Number of Tables / Figures: 1 / 3

Supplementary Materials: 16 pages, pdf

Address for Correspondence

Jonathan Brennan
Linguistics Department, University of Michigan
440 Lorch Hall, 611 Tappan St.
Ann Arbor, MI 48103

jobrenn@umich.edu
734-764-8692

Abstract

Research investigating the brain basis of language comprehension has associated the left anterior temporal lobe with sentence-level combinatorics. Using magnetoencephalography, we test the parsing strategy implemented in this brain region. The number of incremental parse steps from a predictive left-corner parsing strategy that is supported by psycholinguistic research is compared with those from a less-predictive strategy. We test for a correlation between parse steps and source-localized MEG activity recorded while participants read a story. Left-corner parse steps correlated with activity in the left anterior temporal lobe around 350-500 ms after word onset. No other correlations specific to sentence-comprehension were observed. These data indicate that the left anterior temporal lobe engages in combinatoric processing that is well characterized by a predictive left-corner parsing strategy.

1. Introduction

Previous research has implicated the left anterior temporal lobe (ATL) in computing basic aspects of sentence structure during language comprehension (for reviews, see Friederici & Gierhan, 2013; Pylkkänen, 2016). However, the specific algorithm implemented in this region has not been investigated. This study seeks to characterize this algorithm by comparing a predictive "left-corner" parsing strategy and a less-predictive strategy in terms of their fit with the spatio-temporal profile of neural signals recorded using magnetoencephalography during a naturalistic reading task.

Evidence connecting the left ATL to basic sentence combinatorics comes from patient studies, neuroimaging, and electrophysiology. Dronkers et al. (2004) report an analysis of 64 left hemisphere stroke patients who performed a picture-matching task with spoken stimuli that varied in linguistic complexity. Correlations between lesion site and task performance suggested that damage to the left anterior temporal lobe led to difficulty "at the most basic levels of constituent structure processing" (p. 161). Further evidence comes from neuroimaging studies comparing stimuli that do and do not contain sentence structure. Using Positron Emission Tomography (PET), Stowe et al. (1998) compared lists of content and function words with simple sentences, sentences containing long-distance dependencies and sentences with syntactic ambiguities. While the latter two stimulus types led to activation in the left inferior frontal gyrus (IFG; "Broca's Area"; cf. Just et al., 1996; Stromswold et al., 1996) the comparison of simple sentences with word lists led to activation in the anterior temporal lobes bilaterally. Humphries et al. (2005) report that the right ATL is sensitive to the prosodic contours of sentences, while the left is sensitive to the presence or absence of syntactic structure. A focal ATL effect for sentence structure has been replicated using both auditory and visual stimuli (Vandenberghe et al., 2002; Humphries et al., 2006; Rogalsky & Hickok 2009; Mazoyer et al., 1993), while other studies using similar manipulations report activation in this region as well as others (Pallier et al., 2011; Friederici et al., 2000; Jobard et al., 2007; Snijders et al. 2009; Xu et al., 2005; Brennan & Pylkkänen, 2012).

Studies comparing word lists to sentences have not isolated computations specific to sentence parsing from other aspects of sentence comprehension. Brennan et al. (2012, 2016) focused on parsing computations by correlating the amount of structure created word-by-word according to a set of psycholinguistic models with fMRI activity recorded while participants listened to a story. In both studies, a region of the left ATL positively correlated with measures of constituent structure processing. Further, a series of MEG studies has investigated the localization and timing of neural activity involved in understanding simple two-word phrases, like "red boat", with results showing that across a range of tasks and modalities, phrases but not single words or lists of words elicit left anterior temporal activation beginning between 200 and 300 ms after stimulus onset (Bemis & Pylkkänen, 2011, 2012, 2013a,b; Pylkkänen et al. 2014, Blanco-Elorrieta & Pylkkänen, 2015; Zhang & Pylkkänen; Westerlund et al., 2015; Del Prato & Pylkkänen, 2014; Westerlund & Pylkkänen, 2014).

In sum, stimuli that contain phrasal or sentence structure elicit ATL activity (Bemis & Pylkkänen 2011, 2012, 2013a,b; Humphries et al., 2006; Rogalsky & Hickok 2009; Pallier et al., 2011; Friederici et al., 2000; Jobard et al., 2007; Snijders et al. 2009; Vandenberghe et al., 2002; Xu et al., 2005; Brennan & Pylkkänen., 2012, Brennan et al., 2012, 2016) and damage to this region leads to impairments in comprehending simple sentences (Dronkers et al., 2004). This pattern of findings has led to the hypothesis that this region is involved in basic combinatoric processing (Dronkers et al., 2004; Hickok & Poeppel, 2007; Friederici & Gierhan, 2013). Within this converging view, there remains debate as to whether such processes are best characterized in terms of syntactic or semantic composition (Pylkkänen, 2016; see Supplementary Materials for further discussion). In addition, while the literature furnishes a specific hypothesis about basic composition in the left ATL, motivating our focus on it, we also explore possible contributions from other regions that have been implicated in sentence-level computations. These regions include the left temporoparietal junction (TPJ; Humphries et al. 2006; Pallier et al. 2011; Bemis & Pylkkänen 2013), the ventral medial pre-frontal cortex (VMPFC; Pylkkänen & McElree 2007; Brennan & Pylkkänen 2008, 2010; Pylkkänen et al., 2008), and two sub-parts of the left IFG, the Pars Triangularis (PTr) and Pars Opercularis (POp) (e.g. Hagoort, 2005, 2013) (see Supplementary Materials for discussion).

Our goal was to advance our understanding of the algorithm implemented in this sentence processing network, especially the ATL. Sentence processing input is incremental, word-by-word (e.g. Marslen-Wilson 1975; Tanenhaus et al., 1995; Altmann & Kamide, 1999) and research in computational psycholinguistics has identified many candidate parsing strategies that operate in this way (see Hale 2014 for an introduction). These strategies differ in the degree to which they are predictive. At one end of the spectrum are “bottom-up” strategies that posit constituents only when all members of that constituent have been encountered. At the opposite end are “top-down” strategies that predictively postulate structure. Also possible are mixed strategies that postulate a constituent after the left-most member of that constituent has been encountered. Restated in phrase-structure terminology, this “left-corner” strategy identifies each syntactic node immediately after the first child of that node is encountered, but before any remaining children are encountered. Other strategies are available as well, including those that employ different degrees of predictiveness for different grammatical rules (Demers, 1977; Hale 2011), and those that do not fully articulate the constituent structure of a sentence (Sanford & Sturt 2002; Ferreira & Patson, 2007).

One piece of evidence that the left-corner strategy best approximates human performance comes from the memory demands that are imposed by different sentence structures (Abney & Johnson, 1991; Resnik, 1992; Johnson-Laird, 1983). It is well known that humans have trouble processing certain sentences in which one phrase is embedded in the middle of another. For example, the sentence in (1a) can be expanded with a relative clause as in (1b). Continuing this same pattern further, however, leads to comprehension difficulties (1c; Miller and Chomsky, 1963).

(1) a. The plumber visited the house.
b. The plumber [who the contractor likes] visited the house.
c. the plumber [who the contractor [who the homeowner likes] hired] visited the house.

This pattern follows if these "multiply center-embedded" sentences over-tax working memory resources (Miller & Chomsky 1963). Abney & Johnson (1991; cf. Resnick, 1992) show that the left-corner strategy has the property that memory load increases linearly with sentence length for center-embeddings like those in (1) but not for embeddings like those in (2-3) below, which impose linearly increasing memory demands under bottom-up or top-down strategies, respectively. Crucially, humans comprehend these latter types of sentences easily; this pattern follows if the human parser follows a left-corner strategy (cf. Gibson, 1998; Lewis & Vasishth 2005).

(2) Beatrice said that [Susan asked that [Bill tell [Franklin to come home.]]]
(3) [[[Franklin's] Friend's] Sister's] nephew came for a visit.

Numerous other studies have provided evidence that human sentence processing is predictive in a way consistent with the left-corner strategy (Chambers et al., 2002, 2004; Xiang et al., 2009; Sturt & Lombardo 2004; Hale, 2011). Further, the properties of a left-corner strategy are compatible with a broad range of other phenomena from the sentence-processing literature (Hale, 2011, 2014). Given the literature implicating the left ATL for combinatoric processing, we test whether the left-corner strategy characterizes activation in the ATL. We also explore whether it might characterize activation in other sentence-processing regions. Using a context-fee grammar for a fragment of English, we compare a left-corner strategy with a less predictive bottom-up variant in terms of the number of rules evaluated word-by-word. Bottom-up models, which are arguably simpler (Steedman, 2000, p. 230), have shown success in predicting brain signals from the left-anterior temporal lobe recorded with fMRI (Brennan et al., 2012, 2016).

However, fMRI signals are sluggish relative to language processing and are thus not well-suited to adjudicate between different incremental parsing strategies. It is also possible that the parser distinguishes between different kinds of combinatoric rules (e.g. syntactic vs. semantic) such that different rule-types are evaluated by different strategies simultaneously (so-called “asynchronous evaluation” Shieber & Johnson, 1993). We explore this possibility in the Supplementary Materials.

The models that we test are samples from a much larger space of possibilities that ranges over different grammars, strategies for rule evaluation and ambiguity resolution, and hypotheses for linking model dynamics with neural signals (see Brennan, 2016 for discussion). Potential conclusions are constrained by the assumptions that we adopt. Abney and Johnson (1991), for example, use a memory-load metric to link parse strategy with processing cost, while the present study uses a rule-counting metric. Our choices for this initial investigation are guided by recent efforts that successfully link parsing models with behavioral and neural signals using, for example, context-free grammars (e.g. van Schjindel and Schuler, 2015; Henderson et al., 2016), and rule-counting metrics (Brennan et al, 2012; 2016).

We test the model predictions using brain data collected from a story-reading task. Story-reading engages naturalistic processing rather than task-specific strategies that may be elicited by reading isolated sentences. Comprehenders are highly sensitive to the statistical dependencies present in artificial stimuli (Fine et al., 2013), and by using a contemporary short story we seek to minimize idiosyncratic task-specific effects on prediction. A narrative stimulus also increases participant attention and neural engagement (Stephens et al., 2010).

2. Methods

2.1 Participants

27 participants (16 females, 11 males) volunteered for the experiment (age 19 to 33, M = 25). All participants were right-handed (Oldfield, 1971), had normal or corrected-to-normal vision, and reported no history of neurological disorder. All experimental activities were conducted in accordance with the Institutional Review Board at New York University.

2.2 Stimuli and Task

The stimulus was from the short story *Crybaby* by David Sedaris (2008) and was presented visually one word at a time to form the STORY block. Edited for length, the story contained 1279 words. A second LIST block of stimuli was constructed by pseudo-randomizing the story text and removing punctuation to create a list of grammatically unconnected words. The randomization was constrained to avoid pairings of words that that made up possible phrases, including Article-Noun, Adjective-Noun, Preposition-Noun, Noun-Verb and Verb-Noun.

The ordering of the LIST and STORY blocks was counter-balanced across participants. The presentation was interrupted every 1 to 2 minutes by a yes/no memory question probing the content for the STORY block (e.g. “Did the narrator eat lemon chicken for dinner?”) or, for the LIST block, asking about individual words (e.g. “Did you see the word ‘tube’?") or semantic categories (e.g. “did you see any animal names?”). Responses were indicated with a button press on an optical response box placed under the left hand. These questions provided a measure of attention and also offered the participant a short break; the experiment did not resume until initiated by the participant.

2.3 Modeling

Combinatoric rules were applied according to a left-corner strategy. By counting the number of rules evoked word-by-word, the model generated incremental estimates of processing effort (Miller and Chomsky, 1963; Hawkins, 1994; Frazier, 1985); terminal rules were excluded from this calculation.[1] Previous research suggests that the number of rules applied by the parser correlates with associated neural activity (Brennan et al, 2012; 2016).

We used a set of context-free rules defined over syntactic categories. The rules described a grammatical fragment for prepositional phrases (PPs) and we assumed that the parser does not give a label to predicted nodes (see Roark, 2001 pp. 10, 26). In this fragment, a determiner phrase (DP) is the complement of a preposition, and adjectives are adjoined to the noun phrase (NP) that they modify. This constrained domain provided broad enough coverage for the target text while minimizing the number of potentially controversial grammatical assumptions. PP structures are repeated many times throughout the story, thus a detailed parsing analysis of a few examples derived complexity predictions for many phrases with a shared structure. In total, 224

---

[1] More precisely, for a word W and a partial linguistic representation C, where the integration of W with C produces a representation C', the complexity associated with W is the number of rules that have been applied to prove that C' is well-formed.

words in the story fell within the domain of the model. The rules and lexicon for this grammar are given as Supplemental Material and a set of example trees are given in Fig. 1A.

In addition to tracking the left-corner evaluation of syntactic structure, we also tracked the bottom-up evaluation of a set of semantic rules.[2] We have in mind the sense of bottom-up from formal semantics where the interpretation of a constituent is defined in terms of the interpretation of the members of that constituent; this can include sub-constituents whose internal structure has not been entirely recognized (Stabler, 1991). More detail, including the algorithm used, is given in section 3 of the Supplementary Materials.

Examples of the word-by-word rule-application dynamics provided by the model are given in Fig. 1B and Table 1. Each set of counts reifies a hypothesis about the parsing strategy that is implemented in a particular brain region.

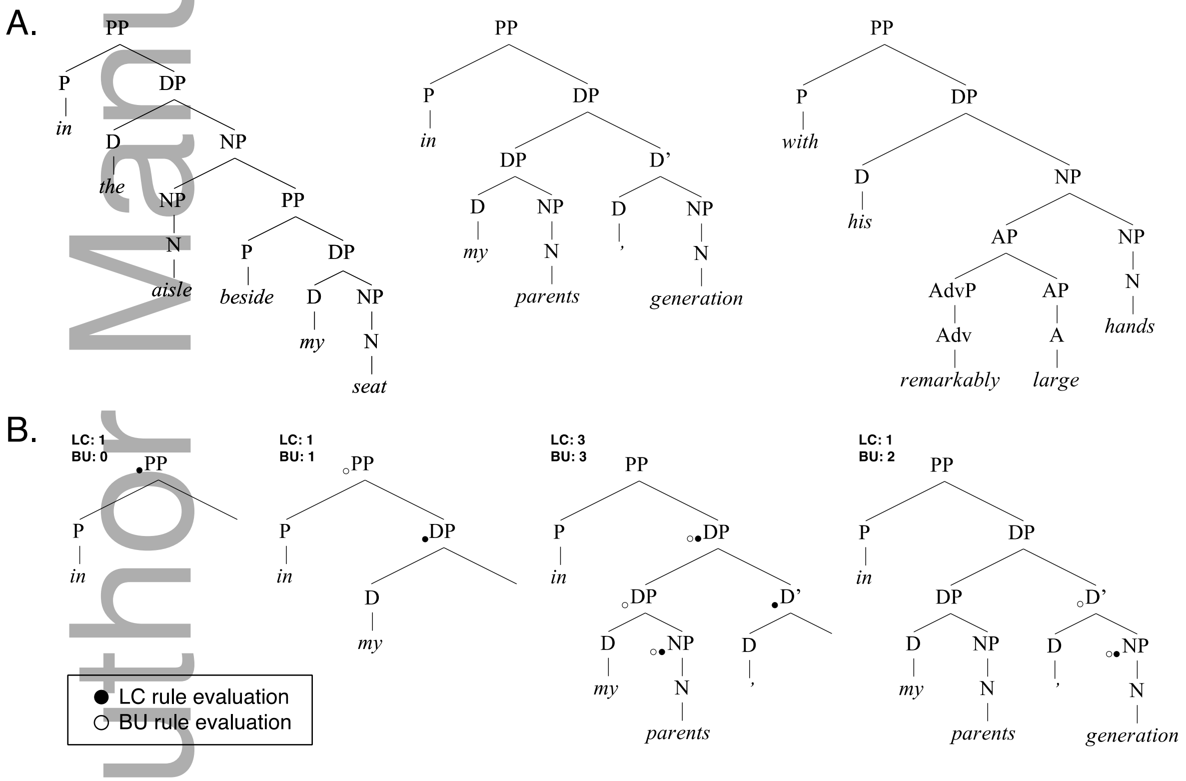


Figure 1. (A) Three example trees for the prepositional phrases covered by the grammar. (B) Word-by-word rule-application dynamics for one example structure. Circles indicate the non-terminal node(s) that are recognized at each step according to the left-corner (LC, closed circle)

---

[2] The rule-sets form a near homomorphism (25 of 27 rules are homomorphic) such that this detail affects only two of the 224 words in the domain of the model.

and bottom-up (BU, open circle) strategies.

Table 1: Example rule-counts from prepositional phrases within the domain of the grammar

1.	*in*	*my*	*parents'*	*generation*		
Left-corner	1	1	3	1		
Bottom-up	0	1	3	2		
2.	*in*	*the*	*aisle*	*beside*	*my*	*seat*
Left-corner	1	1	1	2	1	1
Bottom-up	0	1	2	2	1	2
3.	*with*	*his*	*remarkably*	*large*	*hands*	
Left-corner	1	1	3	1	1	
Bottom-up	0	1	2	2	2	

2.4 MEG data processing

Data were collected in the NYU/KIT MEG facility at New York University. See supplementary materials for details on the experimental procedure.

Data analysis followed that of Brennan & Pylkkänen (2012). Environmental noise recorded at three reference sensors was removed from the data using regression (Adachi et al., 2001). The data were low-pass filtered at 40 Hz, resampled to 200 Hz, and high-pass filtered at 0.1 Hz to remove signal drift. Individual channels showing excessive noise or saturation were excluded (Median = 2, Range = 0–6). Epochs spanning -100 to 600 ms were extracted for all target words in the STORY block that occurred within the domain of the model as were epochs for those same words when they were presented in the LIST block. Three participants were excluded due to recordings with excessive noise, and one participant was excluded due to the lack of characteristic evoked components (i.e. M100, M170; see Pylkkänen and Marantz, 2003). One additional participant was excluded due to a fiducial digitization error, leaving 22 participants for data analysis. Epochs with a peak-to-peak amplitude ≥ 3000 fT were marked as containing an artifact and excluded. On average, 25.1% of STORY epochs and 25.3% of LIST epochs were excluded, leaving an average of 168 and 167 epochs, respectively, per participant. There was no statistical difference in the number of artifacts across the two blocks ($t(21) = 0.13$, $p > 0.5$).

The data were projected into source-space using minimum *l2* norm estimation with MNE software (Martinos Center, MGH, Boston). See Supplementary Materials for source-analysis details.

2.5 Regions of Interest

The analysis focused on a five anatomically constrained regions of interest (ROIs) motivated by prior literature (the region label as it appears in the Freesurfer version 5.1.0 distribution of the Desikan et al. 2006 atlas is given within quotation marks):

(1) Left anterior temporal lobe (ATL; combining the “superiortemporal-lh” and “middletemporal-lh” labels excluding sources posterior to the anterior edge of the transverse temporal gyrus, see Rademacher et al., 1992)
(2) Temporoparietal junction (TPJ; "supramarginal-lh”)
(3) Left pars triangular of the IFG (PTr; "pars-triangularis-lh")
(4) Left pars opercularis of the inferior frontal gyrus (POp; “pars-opercularis-lh")
(5) Left ventromedial prefrontal cortex (VMPFC; “medialorbitofrontal-lh”)

Source time-courses within each region were averaged per epoch. ROI locations are illustrated in Fig. 2.

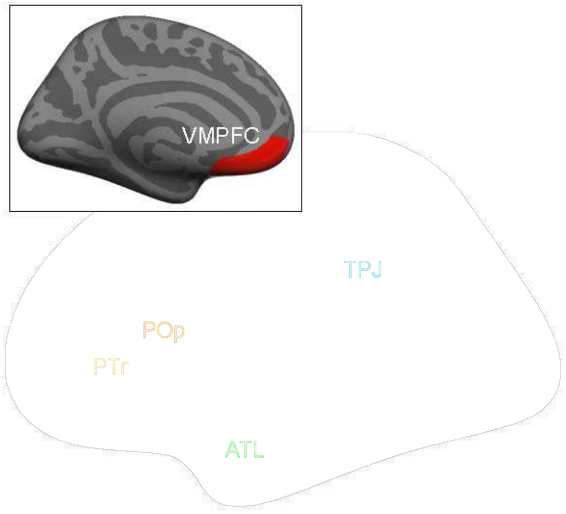


Figure 2: Five regions of interest overlayed on the lateral (insert: medial) inflated cortex of the Freesurfer fsaverage template brain.

2.6 Statistical Analysis

Single-trial source activity and modeled parse steps were compared using linear mixed-effects regression (Gelman & Hill 2006; Baayen et al., 2008). Source estimates per ROI were averaged within 100 ms intervals which spanned -100-600 ms in 50 ms increments. Estimates were modeled as a function of fixed effects for block (STORY or LIST), the number of parse steps estimated by the model, and the interaction between block and parse steps. Models also included nuisance predictors for word length in letters (LEN; mean-centered), trial order (ORD; mean-centered), sentence-position in the story (POS; mean-centered), word frequency (FRQ; log-transformed and mean-centered, based on the HAL written language corpus; Balota et al. 2007), and random intercepts per participant. Parse step and sentence order predictors that were defined based on the STORY block were applied to the same words when the occurred in the LIST block.

The parse step predictors were moderately correlated with several of the nuisance predictors (*r*(Left-corner, FRQ) = -0.26; *r*(Left-corner, LEN) = 0.26; *r*(Bottom-up, FRQ) = -0.53; *r*(Bottom-up, LEN) = 0.46). We thus residualized each parse step predictor against lexical properties of word frequency and word length. We dub the derived coefficients “rLC” and “rBU”. The parse step predictors were also moderately correlated with each other (*r* = 0.39). Steps taken to separately test the independent contributions of these predictors are described in the Supplementary Materials. Correlations between each continuous term entered into the regressions are provided in Supplementary Table S1.

Parse steps may be confounded in some cases with syntactic category: categories more likely to appear towards the beginning of the phrase, such as determiners, are also more likely to have higher scores on a left-corner predictor while categories appearing towards the end of a phrase, such as nouns, are likely to have higher scores on a bottom-up predictor as it is derived from a less-predictive strategy. To ensure effects reflect sentence-level composition and not word-category information, we focused on correlations that were specific to the STORY condition. We did this by conducting a one-tailed test for a positive effect on the block by parse steps interaction coefficient. We determined significance with a non-parametric permutation test in which we created a cluster test statistic by summing standardized coefficients greater than 1.64 ($\alpha = 0.05$ under a normal distribution) from adjacent time-windows. Test statistics were evaluated against a reference distributed created by 10,000 simulations in which we (i) randomly permuted the trial order within participants, (ii) re-fit the regression models against this permuted dependent variable, and (iii) identified the largest cluster statistic per simulation (Maris & Oostenveld, 2007). Cluster statistics with values greater than those from 95% of these simulations were "statistically significant" at $\alpha = 0.05$.

In addition to testing a specific hypothesis about activation in the ATL region, we also explored potential correlations with activation from four other regions. To ensure that our analysis had the same power to detect effects beyond the ATL, we did not impose a different statistical threshold in the exploratory analysis by adding a multiple-comparison correction. Rather, we used the same statistical thresholds across all regions. Any results from the exploratory analysis must, consequentially, be interpreted cautiously.

3. Results

3.1 Behavioral Results

Average accuracy in the STORY block was 88.4% compared with 78.2% in the LIST block. Both of these scores were significantly higher than chance performance (STORY: $t(26) = 21.0$, $p < 0.001$; LIST: $t(26) = 12.1$, $p < 0.001$) and performance in the story block was significantly higher than in the list block, $t(26) = -4.2$, $p < 0.001$.

3.2 MEG Results

We first computed grand-averaged source waves per block for each of the five ROIs. Row 1 of Fig. 3 shows that activation increased over the first 200-400 ms post stimulus onset in all ROIs followed by a decline towards baseline. Peak activation varies slightly across ROIs, with an earlier peak and sustained activation (about 250-400ms) in the ATL contrasting with later peaks around 350-400 ms in the VMPFC, POp and PTr (cf. Brennan & Pylkkänen 2012). TPL activation, in contrast, shows a more subdued response pattern. The significant correlations reported below match the intervals of increased activation within these ROIs.

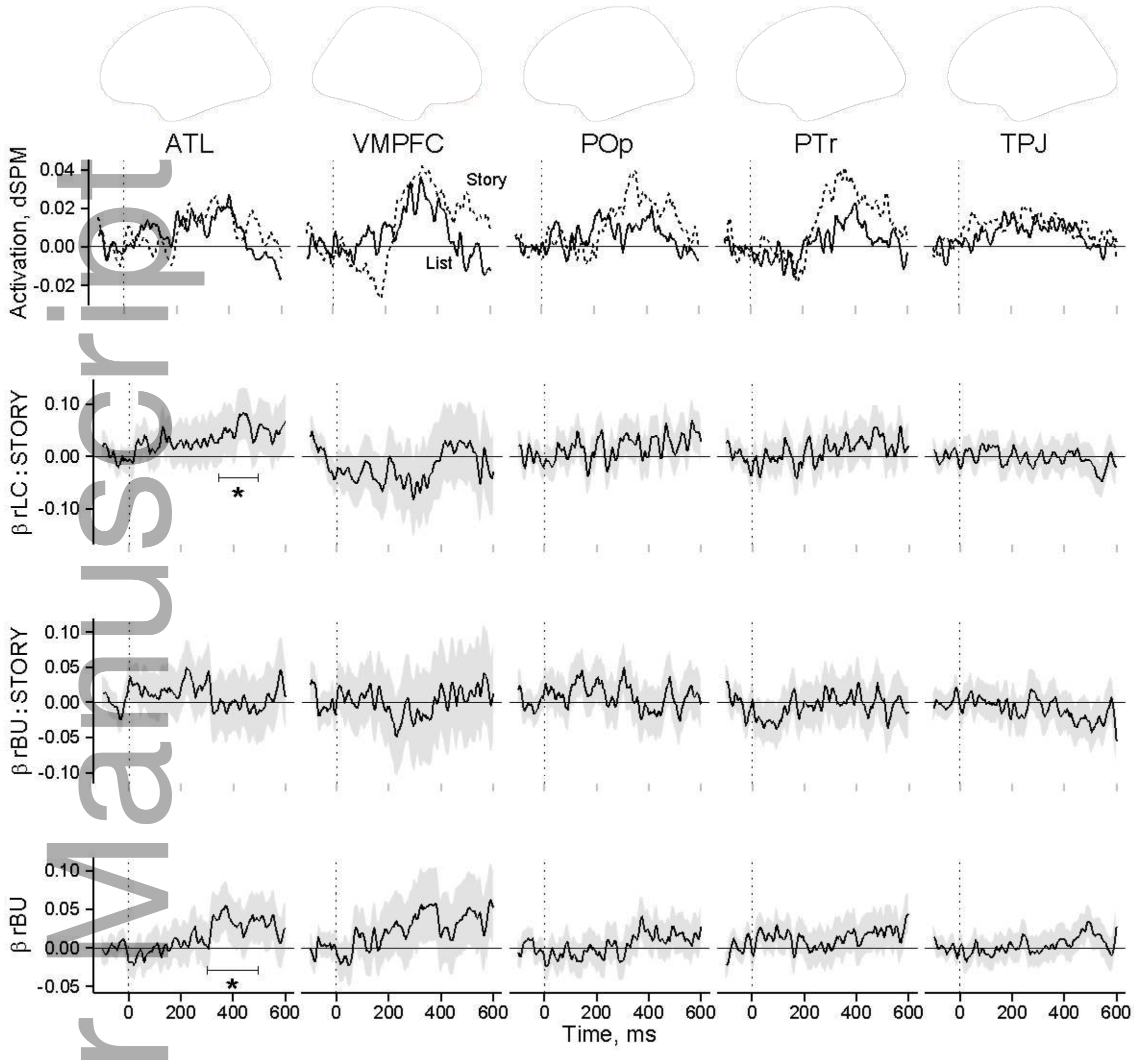


Figure 3: Row 1: Time-courses of averaged source activation from five ROIs for the STORY block (dotted lines) and LIST block (solid). Row 2: Estimated effects (β coefficients) for the interaction of stimulus block with residualized left-corner parse steps (rLC). Row 3: Estimated effects of the interaction of stimulus block with residualized bottom-up parse steps (rBU). Row 4: Estimated main effects of rBU parse steps. Grey shading indicate ±1.64 coefficient standard errors. A positive value for the interaction effects shown in rows 2 and 3 indicates greater effect for left-corner or bottom-up parse steps in the STORY block. ‘*’ indicates a time-span with a statistically significant effect based on a non-parametric permutation test.

The key results are indicated by significant correlations between parse steps (rLC or rBU) that are greater in the STORY block than in the LIST block. This is reflected in an interaction between parse steps and block. Such an interaction was found for rLC in the left ATL from 350-500 ms ($p < 0.05$). Time-courses for the rLC interaction effects are shown in row 2 of Fig. 3. No statistically significant effects were observed in any other region.

No significant effects for the interaction between rBU and block were observed in any ROIs (row 3 of Fig. 3). However, we did observe a significant main effect for rBU in the ATL from 300 to 500 ms ($p < 0.05$; row 4 of Fig. 3). Visual inspection also suggests main effects for rBU in the VMPFC between 300 and 400 ms and a smaller effect in the TPJ between 500 and 600 ms. However, neither of these effects were statistically reliable after correcting for multiple comparisons.

4. Discussion

In this study we recorded magnetoencephalography data during story-reading to test the prediction that ATL-localized brain activity associated with building sentence structure follows a predictive left-corner strategy (Abney & Johnson, 1991; Resnik, 1993; Hale, 2011). Parse steps estimated with the left-corner strategy (rLC) significantly correlated with left ATL brain activity 350-500 ms after stimulus onset for words presented in a story as compared with the same stimuli presented in a randomized list. By residualizing these parse steps against lexical level predictors and by evaluating the contribution of this predictor in the STORY block relative to the LIST block, we sought to isolate the effect for parsing. The concordance between the operations of the left-corner parsing strategy and ATL activity supports the hypothesis that ATL activity may reflect the operations of a circuit that implements this strategy.

The finding that the parse step measure correlates reliably with ATL activity during passive reading of a naturalistic text matches well with previous work showing a correlation between syntactic node count and anterior temporal activity in fMRI when listening to a story (Brennan et al., 2012, 2016). The spatial location is further in accordance with the large body of work showing that anterior temporal lobe is sensitive to the presence of even the simplest phrasal structures, measured with hemodynamic (Humphries et al., 2006; Rogalsky & Hickok 2009; Pallier et al., 2011; Friederici et al., 2000; Jobard et al., 2007; Snijders et al. 2009; Vandenberghe et al., 2002; Xu et al., 2005) and electrophysiological (Bemis & Pylkkänen 2011, 2012, 2013a,b; Brennan & Pylkkänen, 2012; Westerlund & Pylkkänen, 2014; Del Prato & Pylkkänen, 2014; Zhang & Pylkkänen, 2015) techniques. However, the latency of the effect we observe is later than those found in the above-cited electrophysiological studies, which report effects for constituent structure beginning around 250 ms. One speculative explanation for this latency difference is that whereas the studies above examined minimal two-word phrases, the story text used in the present work led to greater variability in the word-by-word time-course of parsing.

We did not observe a significant interaction effect involving parse steps from a less-predictive bottom-up alternative (rBU). We did, however, observe a main effect for rBU: more bottom-up parse steps correlated with increased ATL activity in both the LIST and STORY blocks. This pattern cannot be understood in terms of low-level factors that correlate with word category, such as word frequency or word length, as those confounding variables were factored out using residualization. While any interpretation is necessarily post-hoc, one possibility is that that this correlation reflects word-category information associated with syntactic or semantic frames. Previous work that has focused on word category information by comparing, for example, nouns and verbs, has not found anterior temporal lobe activation (e.g. Bedny & Thompson-Schill, 2006; Berlingeri et al., 2008). However, this work has been conducted using hemodynamic techniques and may have limited sensitivity to the more transient phenomena to which MEG is sensitive.

Recent research has shown that LATL sensitivity to phrasal structure is modulated by the conceptual specificity of the composing lexical items, such that combinatoric effects are obtained only when composition leads to a clear increase in the specificity of the expression (Westerlund & Pylkkänen 2014; Zhang & Pylkkänen, 2015). We did not include a measure of conceptual specificity in our modeling and thus cannot speak to effects of specificity in the current data. However, given the prior specificity findings, our positive LATL results suggest that a sufficient ratio of the combinatoric steps in our narrative satisfied this specificity constraint. Against this backdrop, the interaction of composition and conceptual specificity in narratives is clearly a natural topic for future studies.

We did not find any statistically reliable correlations with left-corner parse steps outside of the ATL region. We interpret such a null result with caution. The present data simply provide no evidence to link incremental parse steps according to the left-corner strategy to regions other than the ATL. Several dimensions of the model remain open for further exploration. One dimension concerns the linking hypothesis. The present approach quantified the cost of moving from one parser state to the next in terms of the number of rules that are evaluated word-by-word. In doing so, the present model assumes a “perfect oracle”: the parser makes the correct choice at each point. Quantifying parser uncertainty (e.g. via “surprisal”; Hale, 2001) offers an alternative linking hypothesis that would provide insight into mechanisms associated with resolving uncertainty (cf. Willems et al., 2015; Henderson et al., 2016; Brennan et al., 2016). Alternatively, one might quantify the memory demands between parser states, for example via memory retrieval effort (Lewis & Vasishth, 2005) or by tracking the depth of the stack for a stack-based parser to tap into mechanisms associated with working memory (Yngve, 1960; Abney & Johnson, 1991).

Another dimension worth exploring concerns the grammar that was used to define well-formed syntactic representations. Alternative analyses of prepositional phrases and noun phrases, including those that permit flexible constituency (Steedman, 2000), or alternatives that vary the hierarchical depth of analysis (e.g. Sanford & Sturt, 2002), are expected to yield estimates distinct from those tested in this experiment. Such data might prove fruitful in testing the predictions of distinct grammatical claims.

5 Conclusion

We tested the prediction that the left anterior temporal lobe implements an operation that can be modeled as left-corner parsing in the service of sentence comprehension. Correlating the number of word-by-word parse steps with MEG data recorded while participants read a story, but not the same words in a random order, revealed increased correlation between parse steps and activity in the left anterior temporal lobe. This result is consistent with and provides algorithmic specificity to the claim that the anterior temporal lobe performs basic combinatoric operations.

Acknowledgements

This work was funded in part by grant G1001 from the NYUAD Institute, New York University Abu Dhabi (L.P.).

References

Abney, S. and Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research*, 20(3):233–250.

Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., and Ochiai, M. (2001). Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. *IEEE Transactions on Applied Superconductivity*, 11(1):669–672.

Allen, K., Pereira, F., Botvinick, M., and Goldberg, A. E. (2012). Distinguishing grammatical constructions with fMRI pattern analysis. *Brain and Language*, 123(3):174–182.

Altmann, G. T. M. and Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3):247–264.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchinson, K. I., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39:445–459.

Bedny, M. and Thompson-Schill, S. L. (2006). Neuroanatomically sep- arable effects of imageability and grammatical class during single-word comprehension. *Brain and Language*, 98(2):127–39.

Bemis, D. K. and Pylkkänen, 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. and Pylkkänen, L. (2012). Combination across domains: An MEG investigation into the relationship between mathematical, pictorial, and linguistic processing. *Frontiers in Psychology*, 3:583.

Bemis, D. K. and Pylkkänen, L. (2013a). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, 23(8):1859–73.

Bemis, D. K. and Pylkkänen, L. (2013b). Flexible composition: Meg evidence for the deployment of basic combinatorial linguistic mechanisms in response to task demands. *PLoS One*, 8(9):e73949.

Berlingeri, M., Crepaldi, D., Roberti, R., Scialfa, G., Luzzatti, C., and Paulesu, E. (2008). Nouns and verbs in the brain: grammatical class and task specific effects as revealed by fMRI. *Cognitive Neuropsychology*, 25(4):528–58.

Brennan, J. (2016). Naturalistic sentence comprehension in the brain. *Language and Linguistics Compass*, 10(7), 299–313.

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., and Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, 120:163–173.

Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W.-M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, 157-158, 81–94.

Brennan, J. and Pylkkänen, L. (2010). Processing psych verbs: Behavioral and MEG measures of two different types of semantic complexity. *Language and Cognitive Processes*, 25(6):777–807.

Brennan, J. and Pylkkänen, L. (2012). The time-course and spatial distribution of brain activity associated with sentence processing. *NeuroImage*, 60:1139–1148.

Chambers, C., Tanenhaus, M., and Magnuson, J. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3):687–696.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., and Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, 47:30–49.

Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, & Computers*, 25(2):257–271.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., and Halgren, E. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, 26(1):55–67.

Del Prato, P. and Pylkkänen, 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.

Demers, A. (1977). Generalized left corner parsing. In *Proceedings of the 4th ACM SIGACT-SIGPLAN symposium on Principles of programming languages*, pages 170–182. ACM New York, NY, USA.

Desikan, R. S., S´egonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., and Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–80.

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., and Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension: Towards a new functional anatomy of language. *Cognition*, 92(1-2):145–177.

Embick, D., Hackl, M., Schaeffer, J., Kelepir, M., and Marantz, A. (2001). A magnetoencephalographic component whose latency reflects lexical frequency. *Cognitive Brain Research*, 10(3):345–348.

Ferreira, F. and Patson, N. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, 1(1–2):71–83.

Fine, A. B., Jaeger, T. F., Farmer, T. A., and Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS One*, 8(10):e77661.

Frazier, L. (1985). Syntactic complexity. In Dowty, D., Karttunen, L., and Zwicky, A. M., editors, *Natural language parsing: Psychological, computational, and theoretical perspectives*, pages 129– 187. Cambridge Univ Press.

Friederici, A. D. and Gierhan, S. M. E. (2013). The language network. *Current Opinion in Neurobiology*, 23(2):250–4.

Friederici, A. D., Meyer, M., and von Cramon, D. Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language*, 74(2):289–300.

Gelman, A. and Hill, J. (2006). *Data analysis using regression and multi- level/hierarchical models*. Cambridge: Cambridge University Press.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.

Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, 23(01):1–21.

Grodzinsky, Y. and Friederici, A. D. (2006). Neuroimaging of syntax and syntactic processing. Current Opinion in Neurobiology, 16(2):240–246.

Hagoort, P. (2005). On broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9(9):416–423.

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, 4(416).

Hagoort, P., Hald, L., Bastiaansen, M., and Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669):438–441.

Hagoort, P. and Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience*, 37:347–62.

Hale, J. (2011). What a rational parser would do. *Cognitive Science*, 35(3):399–443.

Hale, J. T. (2014). *Automaton Theories of Human Sentence Comprehension*. CSLI Publications.

Hale, J. (In press). Information-theoretical complexity metrics. *Language and Linguistics Compass*.

Hämäläinen, M. S. and Sarvas, J. (1989). Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. *IEEE Transactions in Biomedical Engineering*, 36(2):165–71.

Hashimoto, R. and Sakai, K. L. (2002). Specialization in the left prefrontal cortex for sentence comprehension. *Neuron*, 35(3):589–97.

Hawkins, J. (1994). *A Performance Theory of Order and Constituency*. Cambridge University Press.

Henderson, J. M., Choi, W., Lowder, M. W., & Ferreira, F. (2016). Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in reading. *NeuroImage*, 132, 293–300.

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.

Humphries, C., Binder, J. R., Medler, D. A., and Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, 18(4):665–679.

Jobard, G., Vigneau, M., Mazoyer, B., and Tzourio-Mazoyer, N. (2007). Impact of modality and linguistic complexity during reading and listening tasks. *NeuroImage*, 34(2):784–800.

Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge MA: Harvard University Press.

Just, M., Carpenter, P., Keller, T., Eddy, W., and Thulborn, K. (1996). Brain activation modulated by sentence comprehension. *Science*, 274(5284):114–116.

Kamp, H. and Partee, B. (1995). Prototype theory and compositionality. *Cognition*, 57(2):129–191.

Lewis, R. & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.

Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, 189(4198):226–228.

Miller, G. and Chomsky, N. (1963). Finitary models of language users. *Handbook of Mathematical Psychology*.

Oldfield, R. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1):97–113.

Pallier, C., Devauchelle, A.-D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, 108(6):2522–2527.

Pylkkänen, L. (2016). Composition of complex meaning: Interdisciplinary perspectives on the left anterior temporal lobe. In G. Hickok & S. Small (Eds.), *Neurobiology of Language*. Elsevier.

Pylkkänen, L., Brennan, J., and Bemis, D. K. (2011). Grounding the cognitive neuroscience of semantics in linguistic theory. *Language and Cognitive Processes*, 26(9):1317–1337.

Pylkkänen, L. and Marantz, A. (2003). Tracking the time course of word recognition in MEG. *Trends in Cognitive Science*, 7(5):187–189.

Pylkkänen, L., Martin, A. E., McElree, B., and Smart, A. (2008). The anterior midline field: Coercion or decision making? *Brain and Language*, 108(3):184–190.

Pylkkänen, L. and McElree, B. (2007). An MEG study of silent meaning. *Journal of Cognitive Neuroscience*, 19(11):1905–1921.

Pylkkänen, L., Oliveri, B., and Smart, A. (2009). Semantics vs. world knowledge in prefrontal cortex. *Language and Cognitive Processes*, 24:1313–1334.

Rademacher, J., Galaburda, A. M., Kennedy, D. N., Filipek, P. A., and Caviness, Jr, V. S. (1992). Human cerebral cortex: Localization, parcellation, and morphometry with magnetic resonance imaging. *Journal of Cognitive Neuroscience*, 4(4):352–74.

Resnik, P. (1992). Left-corner parsing and psychological plausibility. In *Proceedings of COLING 1992*.

Roark, B. (2001). *Robust probabilistic predictive syntactic processing*. Unpublished PhD dissertation, Brown University.

Rogalsky, C. and Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex*, 19(4):786–796.

Sanford, A. and Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences*, 6(9):382.

Sedaris, S. (2008). *When you are engulfed in flames*. Little, Brown and Company.

Shieber, S. and Johnson, M. (1993). Variations on incremental interpretation. *Journal of Psycholinguistic Research*, 22(2):287–318.

Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J. A., Petersson, K. M., and Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex*, 19(7):1493–503.

Stabler, E. (1991). Avoid the pedestrian's paradox. In Berwick, R. C., Abney, S. P., and Tenny, C., editors, *Principle-Based Parsing: Computation and Psycholinguistics*, pages 199–237. Kluwer Academic Publishers.

Steedman, M. (2000). *The Syntactic Process*. MIT Press.

Stephens, G. J., Silbert, L. J., and Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32):14425–30.

Stowe, L. A., Broere, C. A., Paans, A. M., Wijers, A. A., Mulder, G., Vaalburg, W., and Zwarts, F. (1998). Localizing components of a complex task: Sentence processing and working memory. *Neuroreport*, 9(13):2995–2999.

Stromswold, K., Caplan, D., Alpert, N., and Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language*, 52:452–473.

Sturt, P. and Lombardo, V. (2005). Processing coordinated structures: Incrementality and connectedness. *Cognitive Science*, 29(2):291–305.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational intelligence and neuroscience*, 2011:879716.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., and Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217):1632–1634.

van Schijndel, M. & Schuler, W. (2015). Hierarchic syntax improves reading time prediction. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL 2015).*

Vandenberghe, R., Nobre, A. C., and Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*, 14(4):550–560.

Westerlund, M. and Pylkk ̈anen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia*, 57:59–70.

Wilson, S. M., DeMarco, A. T., Henry, M. L., Gesierich, B., Babiak, M., Mandelli, M. L., Miller, B. L., and Gorno-Tempini, M. L. (2014). What role does the anterior temporal lobe play in sentence-level processing? Neural correlates of syntactic processing in semantic variant primary progressive aphasia. *Journal of Cognitive Neuroscience*, 26(5):970–85.

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & van den Bosch, A. (2015). Prediction during natural language comprehension. *Cerebral Cortex*

Xiang, M., Dillon, B., and Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, 108(1):40–55.

Xu, J., Kemeny, S., Park, G., Frattali, C., and Braun, A. (2005). Language in context: Emergent features of word, sentence, and narrative comprehension. *NeuroImage*, 25(3):1002–1015.

Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society,* 104(5), 444–466.

Zhang, L. and Pylkkänen, L. (2015). The interplay of composition and concept specificity in the left anterior temporal lobe: An MEG study. *NeuroImage*, 111:228–40.

Table Captions

Table 1 Examples of rule counts from prepositional phrases within the domain of grammar when rules are applied according to either a left-corner or bottom-up parsing strategy.

Figure Captions

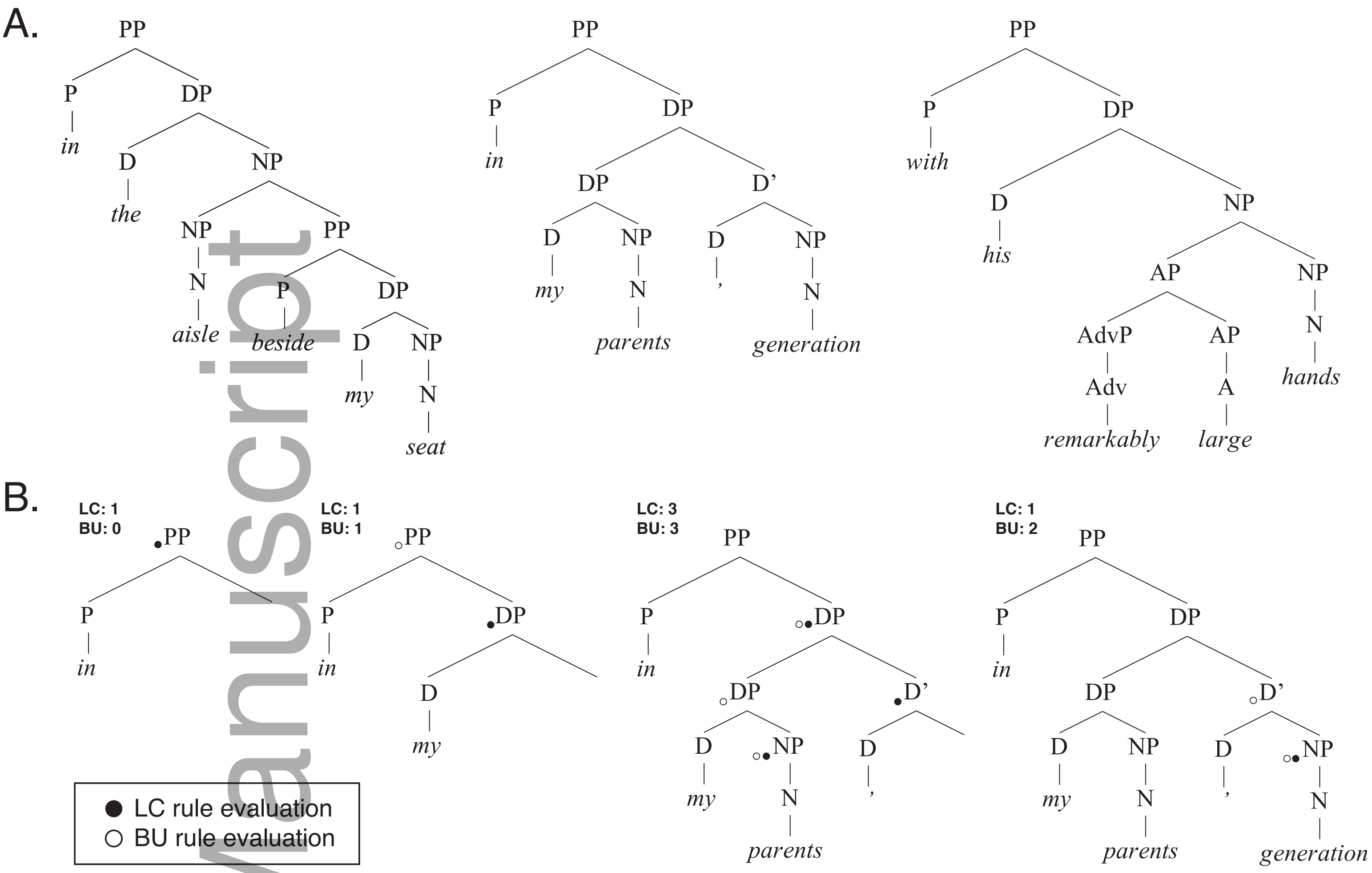
Figure 1. (A) Three example trees for the prepositional phrases covered by the grammar. (B) Word-by-word rule-application dynamics for one example structure. Circles indicate the non-terminal node(s) that are recognized at each step according to the left-corner (LC, closed circle) and bottom-up (BU, open circle) strategies.

Figure 2: Five regions of interest overlayed on the lateral (insert: medial) inflated cortex of the Freesurfer fsaverage template brain.

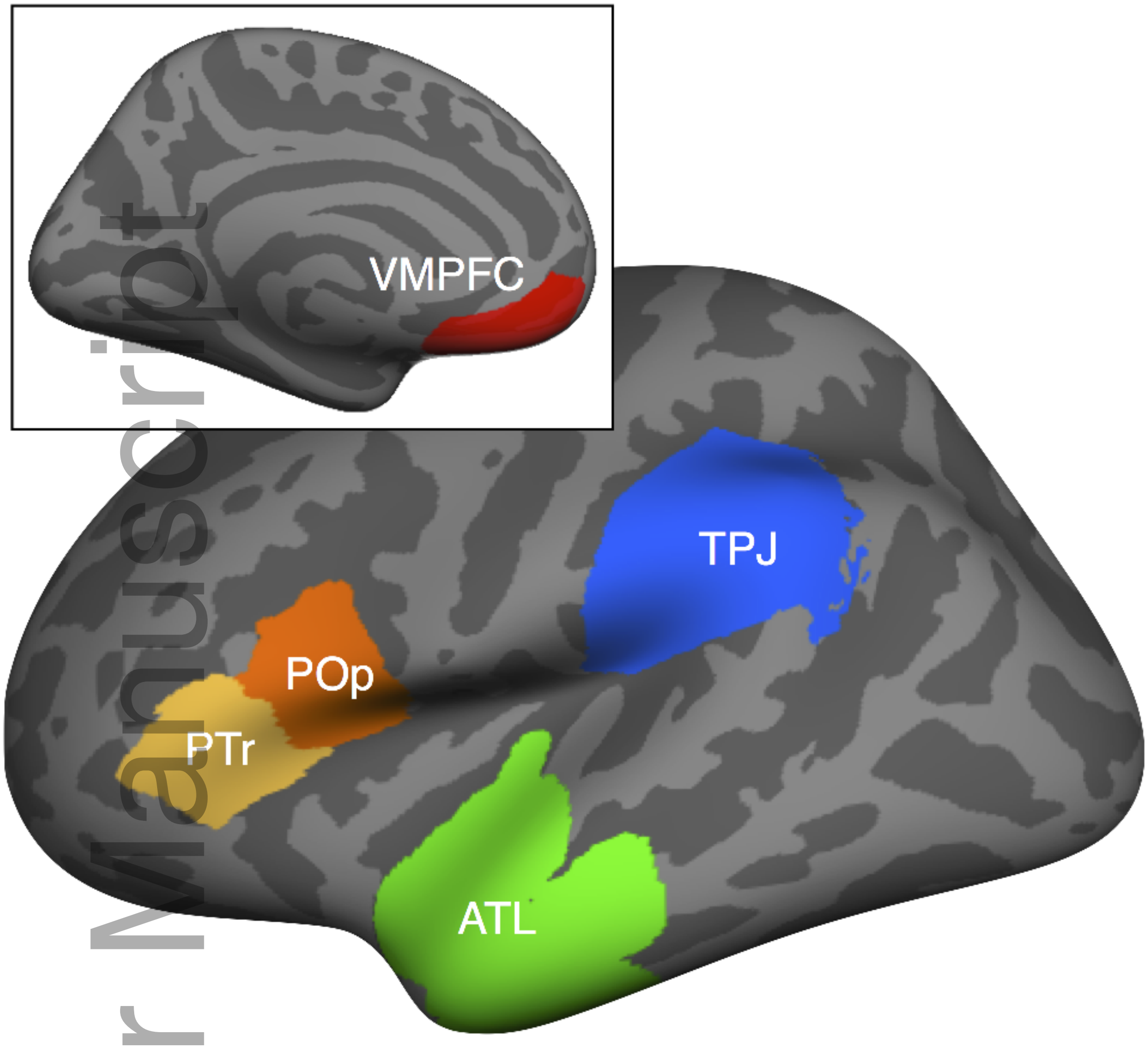
Figure 3: Row 1: Time-courses of averaged source activation from five ROIs for the STORY block (dotted lines) and LIST block (solid). Row 2: Estimated effects (β coefficients) for the interaction of stimulus block with residualized left-corner parse steps (rLC). Row 3: Estimated effects of the interaction of stimulus block with residualized bottom-up parse steps (rBU). Row 4: Estimated main effects of rBU parse steps. Grey shading indicate ±1.64 coefficient standard errors. A positive value for the interaction effects shown in rows 2 and 3 indicates greater effect for left-corner or bottom-up parse steps in the STORY block. ‘*’ indicates a time-span with a statistically significant effect based on a non-parametric permutation test.

Table 1: Example rule-counts from prepositional phrases within the domain of the grammar

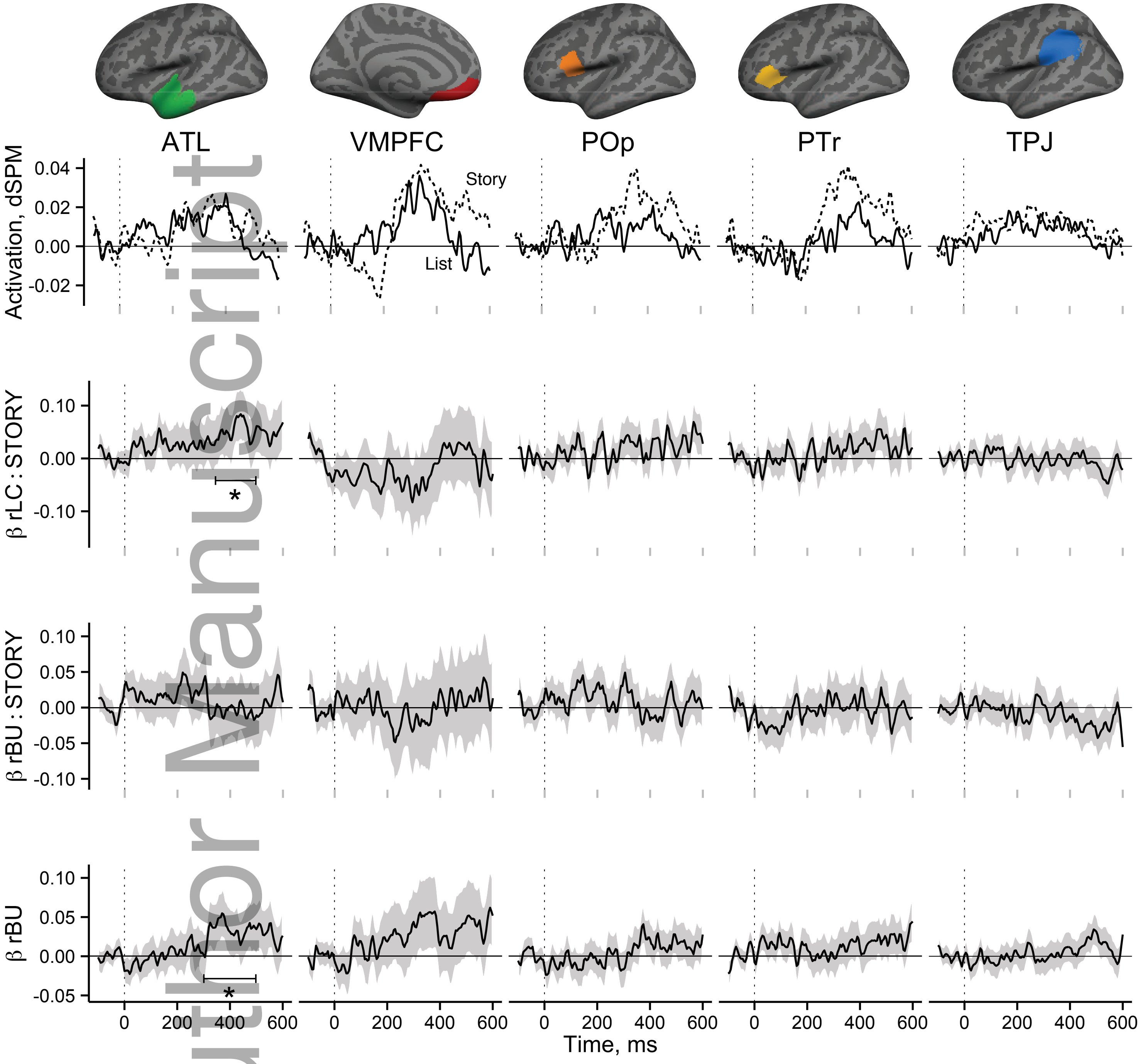
1.	in	my	*parents’*	generation		
Left-corner	1	1	3	1		
Bottom-up	0	1	3	2		
2.	in	the	aisle	beside	my	seat
Left-corner	1	1	1	2	1	1
Bottom-up	0	1	2	2	1	2
3.	with	his	remarkably	large	hands	
Left-corner	1	1	3	1	1	
Bottom-up	0	1	2	2	2	



cogs_12445_f1.eps



cogs_12445_f2.tiff



cogs_12445_f3.eps