COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 41 (2017, Suppl. 6) 1515–1531

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12445

MEG Evidence for Incremental Sentence Composition in the Anterior Temporal Lobe

Jonathan R. Brennan,[a] Liina Pylkkänen[b,c,d]

[a]*Department of Linguistics, University of Michigan*
[b]*Department of Linguistics, New York University*
[c]*Department of Psychology, New York University*
[d]*NYUAD Institute, New York University Abu Dhabi*

Received 20 November 2015; received in revised form 27 July 2016; accepted 12 August 2016

Abstract

Research investigating the brain basis of language comprehension has associated the left anterior temporal lobe (ATL) with sentence-level combinatorics. Using magnetoencephalography (MEG), 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 ATL around 350–500 ms after word onset. No other correlations specific to sentence comprehension were observed. These data indicate that the left ATL engages in combinatoric processing that is well characterized by a predictive left-corner parsing strategy.

Keywords: Syntax; Semantics; Language understanding; Magnetoencephalography

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 spatiotemporal profile of neural signals recorded using magnetoencephalography (MEG) during a naturalistic reading task.

Correspondence should be sent to Jonathan Brennan, Department of Linguistics, University of Michigan, 440 Lorch Hall, 611 Tappan St., Ann Arbor, MI 48109. E-mail: jobrenn@umich.edu

Evidence connecting the left ATL to basic sentence combinatorics comes from patient studies, neuroimaging, and electrophysiology. Dronkers, Wilkins, Van Valin, Redfern, and Jaeger (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 ATL 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, Carpenter, Keller, Eddy, & Thulborn, 1996; Stromswold, Caplan, Alpert, & Rauch, 1996), the comparison of simple sentences with word lists led to activation in the ATLs bilaterally. Humphries, Love, Swinney, and Hickok (2005) report that the right ATL is sensitive to the prosodic contours of sentences, whereas 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 (Humphries, Binder, Medler, & Liebenthal, 2006; Mazoyer et al., 1993; Rogalsky & Hickok, 2009; Vandenberghe, Nobre, & Price, 2002), whereas other studies using similar manipulations report activation in this region as well as others (Brennan & Pylkkänen, 2012; Friederici, Meyer, & von Cramon, 2000; Jobard, Vigneau, Mazoyer, & Tzourio-Mazoyer, 2007; Pallier, Devauchelle, & Dehaene, 2011; Snijders et al., 2009; Xu, Kemeny, Park, Frattali, & Braun, 2005).

Studies comparing word lists to sentences have not isolated computations specific to sentence parsing from other aspects of sentence comprehension. Brennan et al. (2012) and Brennan, Stabler, Van Wagenen, Luh, and Hale (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. Furthermore, 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; Blanco-Elorrieta & Pylkkänen, 2015; Del Prato & Pylkkänen, 2014; Pylkkänen et al. 2014; Westerlund & Pylkkänen, 2014; Westerlund et al., 2015; Zhang & Pylkkänen, 2015).

In sum, stimuli that contain phrasal or sentence structure elicit ATL activity (Bemis & Pylkkänen, 2011, 2012, 2013a,b; Brennan et al., 2012, 2016; Brennan & Pylkkänen, 2012; Friederici et al., 2000; Humphries et al., 2006; Jobard et al., 2007; Pallier et al., 2011; Rogalsky & Hickok, 2009; Snijders et al., 2009; Vandenberghe et al., 2002; Xu et al., 2005) 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; Friederici & Gierhan, 2013; Hickok & Poeppel, 2007). Within this converging view, there remains debate as to whether such processes are best characterized in terms of syntactic or semantic composition (Pylkkänen, 2016; Wilson et al., 2014; 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; Bemis & Pylkkänen, 2013a; Humphries et al., 2006; Pallier et al., 2011), the ventral medial pre-frontal cortex (VMPFC; Brennan & Pylkkänen, 2008, 2010; Pylkkänen & McElree, 2007; Pylkkänen, Martin, McElree, & Smart, 2008), and two subparts 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., Altmann & Kamide, 1999; Marslen-Wilson, 1975; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), 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 (Ferreira & Patson, 2007; Sanford & Sturt, 2002).

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; Johnson-Laird, 1983; Resnik, 1992). 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 & 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 and Johnson (1991; cf. Resnik,

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, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Chambers, Tanenhaus, & Magnuson, 2004; Hale, 2011; Sturt & Lombardo, 2005; Xiang, Dillon, & Phillips, 2009). Furthermore, 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-free 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-ATL 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, whereas this 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., Henderson, Choi, Lowder, & Ferreira, 2016; van Schijndel & Schuler, 2015) 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, Jaeger, Farmer, & Qian, 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, Silbert, & Hasson, 2010).

2. Methods

2.1. Participants

A total of 27 participants (16 women, 11 men) volunteered for the experiment (age 19–33, $M = 25$). 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 1,279 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 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 counterbalanced across participants. The presentation was interrupted every 1–2 min 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 (Frazier, 1985; Hawkins, 1994; Miller & Chomsky, 1963); 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 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 words in the story fell within the domain of the model. The rules and lexicon for this grammar are given in Supplementary 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 subconstituents 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.

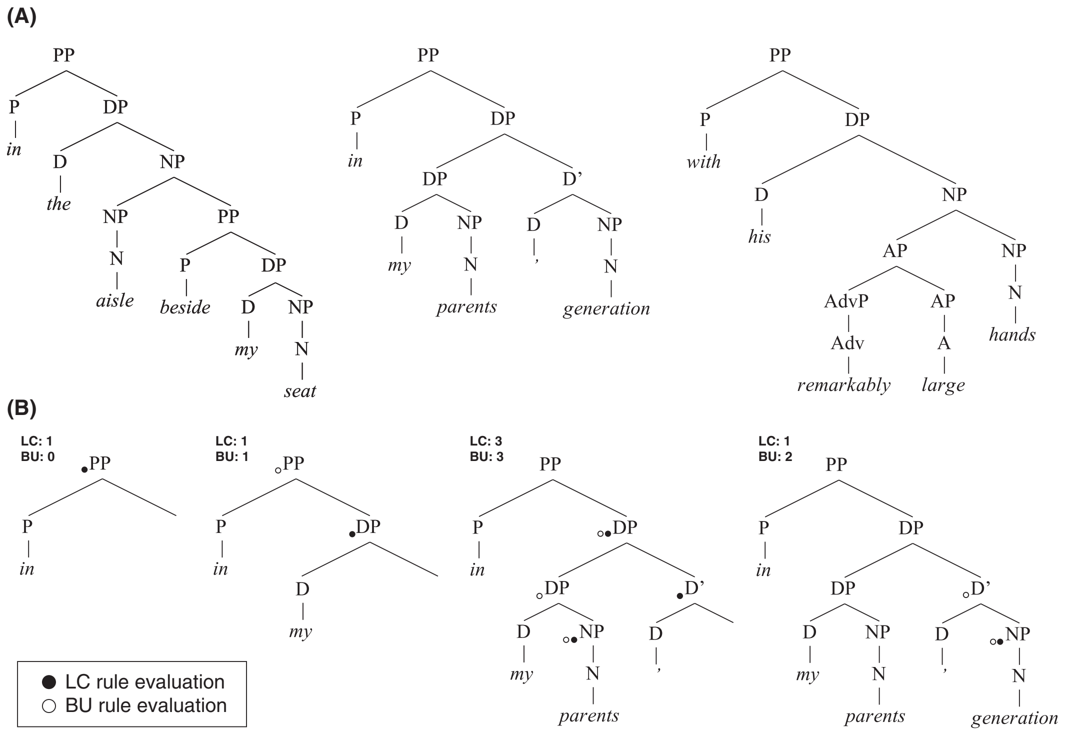

Fig. 1. (A) Three example trees for the prepositional phrases (PPs) covered by the grammar. (B) Word-by-word rule application dynamics for one example structure. Circles indicate the non-terminal node(s) that are recognized at each step according to the left-corner (LC, closed circle) and bottom-up (BU,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

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

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 and Pylkkänen (2012). Environmental noise recorded at three reference sensors was removed from the data using regression (Adachi, Shimogawara, Higuchi, Haruta, & Ochiai, 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 & 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 ≥3,000 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, Galaburda, Kennedy, Filipek, & Caviness, 1992)
2. Temporoparietal junction (TPJ; “supramarginal-lh”)
3. Left pars triangular of the inferior frontal gyrus (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.

2.6. Statistical analysis

Single-trial source activity and modeled parse steps were compared using linear mixed-effects regression (Baayen, Davidson, & Bates, 2008; Gelman & Hill, 2006). Source estimates per ROI were averaged within 100 ms intervals which spanned -100 to 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 they occurred in the LIST block.

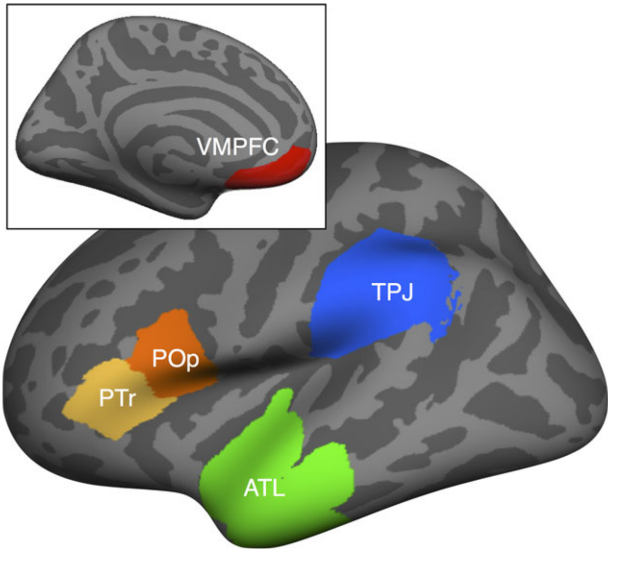

Fig. 2. Five regions of interest overlayed on the lateral (insert: medial) inflated cortex of the Freesurfer fsaverage template brain.

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 numeric 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 toward the beginning of the phrase, such as determiners, are also more likely to have higher scores on a left-corner predictor, whereas categories appearing toward 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 >1.64 (α = 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 (a) randomly permuted the trial order within participants, (b) re-fit the regression models against this permuted dependent variable, and (c) 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 α = 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 toward baseline. Peak activation varies slightly across ROIs, with an earlier peak and sustained activation (about 250–400 ms) in the

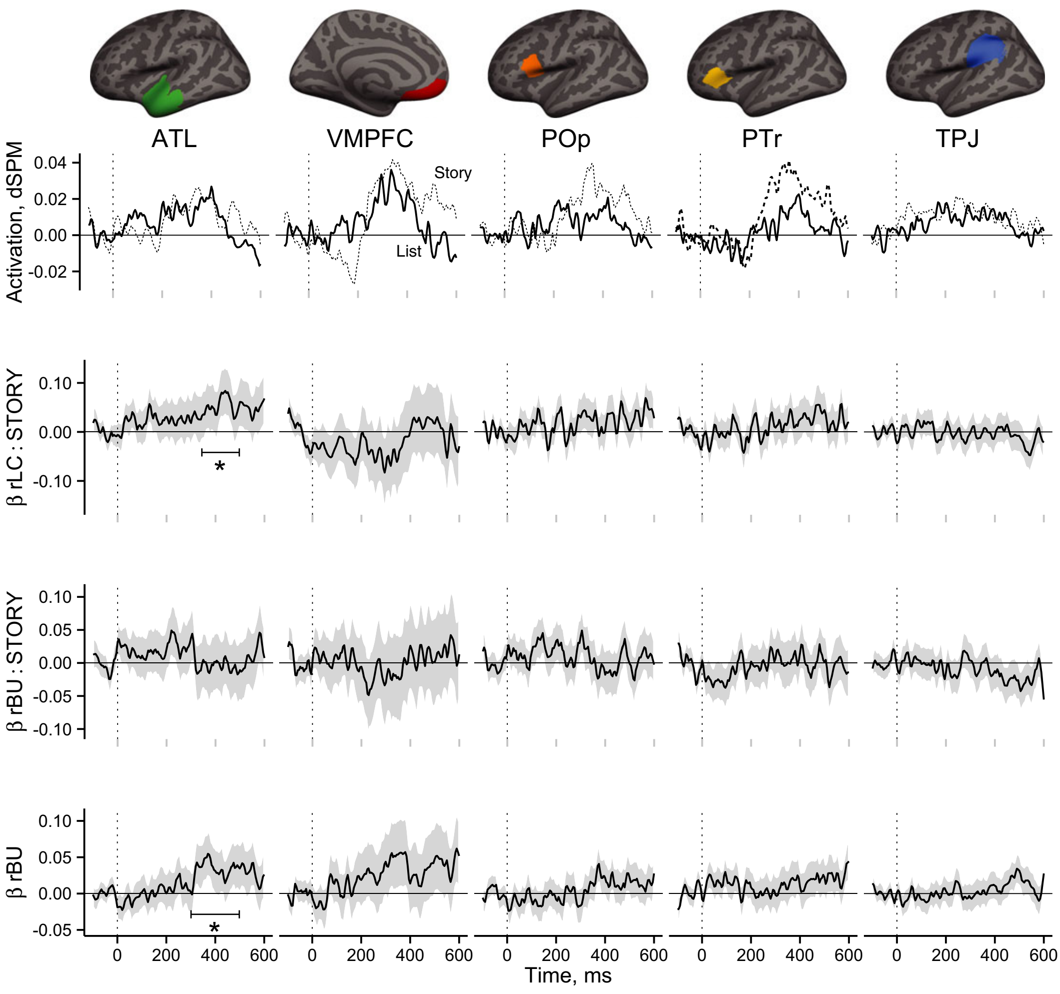

Fig. 3. Row 1: Time courses of averaged source activation from five ROIs for the STORY block (dotted lines) and LIST block (solid). Row 2: Estimated effects (β coefficients) for the interaction of stimulus block with residualized left-corner parse steps (rLC). Row 3: Estimated effects of the interaction of stimulus block with residualized bottom-up parse steps (rBU). Row 4: Estimated main effects of rBU parse steps. Gray shading indicates ±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.

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.

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 to 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 across time.

4. Discussion

In this study we recorded MEG 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; Hale, 2011; Resnik, 1992). 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 ATL is sensitive to the presence of even the simplest phrasal structures, measured with hemodynamic (Friederici et al., 2000; Humphries et al., 2006; Jobard et al., 2007; Pallier et al., 2011; Rogalsky & Hickok, 2009; Snijders et al., 2009; Vandenberghe et al., 2002; Xu et al., 2005) and electrophysiological (Bemis & Pylkkänen, 2011, 2012, 2013a,b; Brennan & Pylkkänen, 2012; Del Prato & Pylkkänen, 2014; Westerlund & 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 of this latency difference is that whereas the studies above examined minimal two-word phrases, the story text used in this 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. Although any interpretation is necessarily post hoc, one possibility is 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 ATL 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 these 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. These 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. Our 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. Brennan et al., 2016; Henderson et al., 2016; Willems, Frank, Nijhof, Hagoort, & van den Bosch, 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 (Abney & Johnson, 1991; Yngve, 1960).

Another dimension worth exploring concerns the grammar that was used to define well-formed syntactic representations. Alternative analyses of PPs and NPs, 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 ATL 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 a positive correlation between parse steps and activity in the left ATL. This result is consistent with and provides algorithmic specificity to the claim that the ATL performs basic combinatoric operations.

Acknowledgments

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

Notes

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.
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.

References

Abney, S., & 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., & 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.

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

Baayen, R. H., Davidson, D. J., & 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., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*, 445–459.

Bedny, M., & Thompson-Schill, S. L. (2006). Neuroanatomically separable effects of imageability and grammatical class during single-word comprehension. *Brain and Language*, *98*(2), 127–139.

Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *The Journal of Neuroscience*, *31*(8), 2801–2814.

Bemis, D. K., & 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., & 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), 1859–1873.

Bemis, D. K., & 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., & Paulesu, E. (2008). Nouns and verbs in the brain: Grammatical class and task specific effects as revealed by fMRI. *Cognitive Neuropsychology*, *25*(4), 528–558.

Blanco-Elorrieta, E., & Pylkkänen, L. (2015). Brain bases of language selection: MEG evidence from Arabic-English bilingual language production. *Frontiers in Human Neuroscience*, *9*, 27.

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., & Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, *120*, 163–173.

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

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.

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

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

Del Prato, P., & 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* (pp. 170–182). New York, NY: ACM New York.

Desikan, R. S., Ségonne, 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., & 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), 968–980.

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: Towards a new functional anatomy of language. *Cognition*, *92*(1–2), 145–177.

Ferreira, F., & 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., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8*(10), e77661.

Frazier, L. (1985). Syntactic complexity. In D. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 129–187). Cambridge, UK: Cambridge University Press.

Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. *Current Opinion in Neurobiology*, *23* (2), 250–254.

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

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

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

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).

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In North American (Ed.), *chapter of the Association for Computational Linguistics* (pp. 1–8). Morristown, NJ: Association for Computational Linguistics.

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*. Stanford, CA: CSLI Publications.

Hawkins, J. (1994). *A performance theory of order and constituency*. Cambridge, UK: 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.

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. A., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, *18*(4), 665–679.

Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Human Brain Mapping*, *26*(2), 128–138.

Jobard, G., Vigneau, M., Mazoyer, B., & 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., & Thulborn, K. (1996). Brain activation modulated by sentence comprehension. *Science*, *274*(5284), 114–116.

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

Maris, E., & 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.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., & Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, *5*(4), 467–479.

Miller, G., & Chomsky, N. (1963). Finitary models of language users. In R. R. Bush, E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 419–491). New York: John Wiley & Sons Inc.

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

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.

Pylkkänen, L. (2016). Composition of complex meaning: Interdisciplinary perspectives on the left anterior temporal lobe. In G. Hickok & S. Small (Eds.), *Neurobiology of language* (pp.621–631). London: Elsevier.

Pylkkänen, L., Bemis, D. K., & Blanco Elorrieta, E. (2014). Building phrases in language production: An MEG study of simple composition. *Cognition*, *133*(2), 371–384.

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

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

Rademacher, J., Galaburda, A. M., Kennedy, D. N., Filipek, P. A., & Caviness Jr.,V. S. (1992). Human cerebral cortex: Localization, parcellation, and morphometry with magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *4*(4), 352–374.

Resnik, P. (1992). Left-corner parsing and psychological plausibility. In *Proceedings of the 14th conference on computational linguistics* (pp. 191–197).Stroudsburg, PA: Association for Computational Linguistics.

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

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.

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

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)* (pp. 1597–1605). Association for Computational Linguistics: Stroudsburg, PA.

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

Shieber, S., & 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., & 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. (1991). Avoid the pedestrian's paradox. In R. C. Berwick, S. P. Abney, & C. Tenny (Eds.), *Principle-based parsing: Computation and psycholinguistics* (pp. 199–237). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Steedman, M. (2000). *The syntactic process*. Cambridge, MA: MIT Press.

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*(32), 14425–14430.

Stowe, L. A., Broere, C. A., Paans, A. M., Wijers, A. A., Mulder, G., Vaalburg, W., & 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., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language*, *52*, 452–473.

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

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

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.

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

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & van den Bosch, A. (2016). Prediction during natural language comprehension. *Cerebral Cortex*, *26*(6), 2506–2516.

Wilson, S. M., DeMarco, A. T., Henry, M. L., Gesierich, B., Babiak, M., Mandelli, M. L., Miller, B. L., & 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), 970–985.

Xiang, M., Dillon, B., & 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., & 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., & Pylkkänen, L. (2015). The interplay of composition and concept specificity in the left anterior temporal lobe: An MEG study. *NeuroImage*, *111*, 228–240.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. (1) Region of interest background, (2) Methods: MEG procedure, (3) Auxiliary analysis: Asynchronous evaluation, (4) Grammar and lexicon.