Main Manuscript for

'Constituent length' effects in fMRI do not provide evidence for abstract syntactic processing

*Cory Shain[1,2], Hope Kean[1,2], Benjamin Lipkin[1,2], Josef Affourtit[1,2], Matthew Siegelman[1,2,3], Francis Mollica[4 †], and Evelina Fedorenko[1,2,5 †]

†Co-senior authors

[1]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, USA

[2]McGovern Institute for Brain Research, MIT, Cambridge, MA 02139, USA

[3]Department of Psychology, Columbia University, New York, NY 10027, USA

[4]Department of Informatics, Edinburgh University, Edinburgh, U.K.

[5]The Program in Speech and Hearing in Bioscience and Technology, Harvard University, Boston, MA 02114

*Cory Shain

Email: cshain@mit.edu

PNAS strongly encourages authors to supply an ORCID identifier for each author. Do not include ORCIDs in the manuscript file; individual authors must link their ORCID account to their PNAS account at www.pnascentral.org. For proper authentication, authors must provide their ORCID at submission and are not permitted to add ORCIDs on proofs.

Author Contributions: Conceptualization: CS, FM, EF; Design and materials creation: MS, MF and EF; Experimental script creation: MS; fMRI data collection: HK, JA, MS, EF; fMRI data preprocessing and analysis: all authors; Formal statistical analysis: CS; Figures: CS; Writing original draft: CS and EF; Editing and comments: HK, BL, and FM; Overall supervision: FM and EF.

Competing Interest Statement: The authors have no competing interests to disclose.

Classification: Biological Sciences – Neuroscience ; Social Sciences – Cognitive Sciences

Keywords: language; syntax; constituency; lexicon; abstraction

This PDF file includes:

Main Text
Figures 1 to 2
Tables 1 to 1

Abstract

How are syntactically and semantically connected word sequences, or *constituents*, represented in the human language system? An influential fMRI study, Pallier et al. (2011, *PNAS*), manipulated the length of constituents in sequences of words or pseudowords. They reported that some language regions (in the anterior temporal cortex and near the temporo-parietal junction) were sensitive to constituent length only for sequences of real words but not pseudowords. In contrast, language regions in the inferior frontal and posterior temporal cortex showed the same pattern of increased response to longer constituents—and similar overall response magnitudes—for word and pseudoword sequences. Based on these results, Pallier et al. argued that the latter regions represent abstract sentence structure. Here we identify methodological and theoretical concerns with the Pallier et al. study and conduct a replication across two fMRI experiments. Our results do not support Pallier et al.'s critical claim of distinct neural specialization for abstract syntactic representations. Instead, we find that all language regions show a similar profile of sensitivity to both constituent length and lexicality (stronger responses to real-word than pseudoword stimuli). In addition, we argue that the constituent length effect in these experiments i) is not readily grounded in established theories of sentence processing, and ii) may not actually derive from syntactic structure building, but may instead reflect the temporal receptive window of the human language system.

Significance Statement

An influential study (Pallier et al., 2011, *PNAS*) has been widely interpreted as evidence that inferior frontal and posterior temporal brain regions perform abstract syntactic processing during language comprehension, independent of lexical content. Here we identify theoretical, empirical, and methodological concerns with the original study and conduct a replication across two experiments. Contrary to Pallier et al. (2011), we find evidence that both syntactic and lexical processing are distributed throughout the human language network, consistent with prior arguments for a highly lexicalized and distributed human language processing system.

Main Text

Introduction

Human languages are characterized by rich and complex structure. How sentence structure is processed during real-time comprehension is a central question in the study of language (1). In an influential study, Pallier, Devauchelle, and Dehaene (ref. (2), henceforth *PDD*) provided fMRI evidence that syntactic constituents—groups of words that function as single units within a hierarchical structure of a sentence—are represented in the brain when people read sentences. More interestingly, they argued that their evidence showed that brain regions in the inferior frontal and posterior temporal cortex represent *abstract syntactic structure* without the lexical content of sentences. Ten years later, PDD has been cited over 500 times, and its claims have informed theories of cognition, brain function, and evolution that posit neural circuits dedicated to abstract combinatorics (e.g., refs. (3–7)).

In PDD's paradigm (**Figure 1**), participants read 12-word sequences presented one word at a time. The internal composition of the sequences varied parametrically from a sequence of twelve unconnected words to a 12-word sentence (condition "c12" in **Figure 1**). In the intermediate conditions, the sequences contained concatenated constituents of different lengths: six 2-word constituents (c02), four 3-word constituents (c03), three 4-word constituents (c04), or two 6-word constituents (c06). PDD hypothesized that normal language processing requires the comprehender to maintain an increasingly complex representation of constituent structure as each new word is processed, and that this increased representational complexity will correspond to an increase in overall neuronal activity in conditions with longer constituents. To investigate the abstractness of syntactic representations, a 'Jabberwocky' version of each condition (e.g. jab-c01, jab-c12) was created by replacing the content words (nouns, verbs, adjectives, and adverbs) with word-like nonwords (pseudowords), but preserving the syntactic 'frame', i.e., function words like articles and auxiliaries, and functional morphological endings (e.g., *higher and higher prices* > *hisker and hisker cleeces*).

In line with their hypothesis, PDD observed stronger neural responses to real-word sequences comprised of longer constituents in six frontal and temporal left-hemisphere regions previously associated with language processing. Critically, they found that Jabberwocky sequences elicited a similar-magnitude response and a similar pattern of constituent-length effects in inferior frontal and posterior temporal, but not anterior temporal or temporo-parietal regions, leading to the argument that these regions represent abstract syntactic structure and are insensitive to word meanings.

However, PDD's core claims now face empirical and theoretical objections. First, multiple past studies have found evidence of lexical processing in the inferior frontal and posterior temporal areas identified by PDD as abstract syntactic hubs (e.g., refs. (8–12)), and other studies have reported sensitivity to structure in Jabberwocky materials in anterior temporal regions argued by PDD to be insensitive to such effects (e.g., refs. (8–10, 13, 14)). These prior studies raise concerns about the empirical validity of PDD's reported pattern. Second, PDD's proposed theory of syntactic structure building (which predicts a monotonic increase in demand across the constituent) is at odds with an extensive theoretical and empirical literature on human sentence processing that has revealed considerable variation in processing demand over the course of constituents (15), including *reductions* in demand for certain kinds of long constituents (16, 17).

Furthermore, some of the methodological choices in PDD's design and analyses are problematic. First, PDD used a between-subjects design to compare the real-words and Jabberwocky conditions (thus simultaneously varying both the sample of participants and the condition), even though this manipulation is feasible to perform in a within-subjects design that avoids this confound. Because individuals and, by extension, groups of individuals vary along numerous trait and state dimensions that are known to affect neural responses (e.g., refs. (18–20)), the magnitudes of neural responses in two groups cannot be confidently attributed to differences/similarities between conditions. Second, PDD used the same data both to define the regions of interest and to quantify their responses, introducing circularity (21). Finally, PDD relied on traditional group analyses (18), which assume voxel-wise correspondence across individual brains. Ample evidence exists for substantial inter-individual variability in the precise locations of functional areas in the association cortex (e.g., refs. (22–24)), including in the language network (e.g., refs. (8, 25)). Given that some of PDD's claims rely on not finding certain effects in certain brain regions, the choice of traditional group analyses, which suffer from low sensitivity (26) is suboptimal.

Motivated by these concerns, we conduct two experiments that constitute the closest effort to date to replicate PDD's original study while addressing the methodological issues above. First, we use a strictly within-subjects design. Second, we use independent data to define the regions of interest and to quantify their responses to the critical conditions. And third, we define areas of interest functionally in individual brains (e.g., refs. (8, 27, 28)), which has been shown empirically and through simulations to yield higher sensitivity and higher functional resolution (e.g., refs. (26, 29–31)).

Whereas we replicate the basic constituent-length effect in both experiments (see ref. (32) for another recent replication), our results challenge PDD's critical claim that the inferior frontal and posterior temporal regions support abstract syntactic processing. In particular, all language regions show (a) an effect of 'lexicality' with real-word conditions eliciting stronger responses than Jabberwocky conditions, (b) a length by lexicality interaction whereby the constituent-length effect is more pronounced in the real-word compared to Jabberwocky conditions, or (c) both. These findings challenge the notion of regions within the language network that support abstract, content-independent, syntactic processing.

We further show that multiple extant theories of human language processing do not explain PDD's pattern of results. This finding makes it difficult to ground PDD's effect in independently motivated mechanisms of sentence processing. We propose a non-syntactic alternative account of PDD's constituent-length effect in terms of the size of the language system's temporal receptive window (e.g., ref (33)) that aligns with prior research.

Results

Results are visualized in **Figure 2** (full significance testing details are given in **Table S1**). For the real-word conditions, all regions show the pattern reported by PDD: increasing activation as a function of constituent length, including a smaller increase at larger lengths (e.g., c06 to c12). This pattern is robust in both Experiment 1 and 2 (**Figure 2B-D**). However, as shown in **Figure 2D**, both a) the language network when treated as an integrated whole (see e.g., refs (31, 34–36)), and b) each individual region within it (correcting over regions for false discovery rate – FDR; see **Materials & Methods**) also show i) constituent-length effects for the Jabberwocky conditions (significant for all but the LAngG language fROI), ii) lexicality effects (larger overall responses to real-word than Jabberwocky stimuli; significant for all but the LIFGorb language fROI), and iii) constituent-length by lexicality interactions (larger constituent-length effects for real-word than Jabberwocky conditions; significant in the LIFGorb, LAntTemp, and LAngG language fROIs, and in the language network overall). Thus, contrary to PDD, who reported the same response pattern to real-word and Jabberwocky stimuli in inferior frontal and posterior temporal regions, we find significant effects of stimulus type in these regions, either in the form of larger overall response to real-word stimuli (the LIFG and LPostTemp language fROIs) or steeper increases in response to constituent length in real-word stimuli (the LIFGorb language fROI). In summary, no region exhibits the critical pattern of similar sensitivity to the constituent-length manipulation (which PDD argue is a *syntactic* manipulation, but see **Discussion**) in the absence of sensitivity to lexical content (i.e., real words vs. Jabberwocky).

To help interpret the constituent-length effect observed by PDD and replicated here (see also ref. (32)), follow-up analyses considered the impact of including as predictors in the model six linguistic measures that are motivated by an extensive theoretical and empirical literature on human language processing mechanisms and their cognitive demands: open nodes, node closings, Dependency Locality Theory (DLT) storage cost, DLT integration cost, 5-gram surprisal, and PCFG surprisal (see **Materials & Methods** and **SI Section 5**). If the constituent-length effect is due to one or more of these linguistic variables, then controlling for them should attenuate the effect. However, under the same FDR correction as above, no linguistic variable significantly alters the strength of the overall constituent-length effect in the language network in either experiment. In other words, we find no evidence that PDD's pattern of results can be explained by (or grounded in) prevailing theories of cognitive load during language comprehension. As we argue below, the constituent-length effect may be driven primarily by poorer overall engagement of the language processing system in shorter conditions, rather than by syntactic structure building as argued by PDD.

Discussion

By showing purported evidence for the existence of brain regions specialized for representing abstract linguistic structure PDD provided an important connection between the brain, cognition, and the structure of natural language that has informed much subsequent theorizing about the neural basis of language and the structure of mental representations for language (e.g., refs. (3–7)). However, PDD's conclusions (1) relied on statistically questionable between-group comparisons to substantiate the claim of abstract syntactic processing, (2) used the same data to define the fROIs and to statistically examine their responses, (3) did not take into account individual variation in functional brain anatomy, and (4) depended on a theory of language processing that has not been externally validated, conflicts with known empirical patterns, and is not widely accepted by the sentence processing community. In two conceptual replications that used independent data to define the areas of interest and to quantify their responses (e.g., (21)), we reproduced PDD's finding of increased language network activation as a function of constituent length for real-word stimuli. However, contrary to PDD, we find that (1) no language region shows a pattern consistent with abstract syntactic processing, and (2) all language regions, except for the LAngG language fROI, show qualitatively and, for the most part, quantitatively similar patterns of response, arguing against the division proposed by PDD between abstract syntactic regions (inferior frontal and posterior temporal regions) and regions that are only sensitive to syntactic structure in meaningful stimuli (anterior temporal regions). These key similarities and differences between our findings and PDD's are summarized in **Table 1.**

PDD's core claim is that the *inferior frontal and posterior temporal components of the language network* (but not its anterior temporal or temporo-parietal components) *support abstract syntactic processing* given that—in their data—these regions show similarly strong responses to real-word and Jabberwocky stimuli, and similar constituent-length effects. The similar magnitudes of response to real-word and Jabberwocky stimuli may have been an artifact of a between-group comparison (separate groups of participants performed the real-word conditions vs. the Jabberwocky conditions). Here, using a within-subjects design, we show a robust effect of lexicality such that real-word stimuli elicit a much stronger response than Jabberwocky stimuli. This effect is present across the language network, and critically in both the inferior frontal and posterior temporal areas (the LIFG and LPostTemp language fROIs). This finding aligns with several prior studies (fMRI: ref. (8)—see **Figure S1-B** for a direct comparison of the overlapping subset of conditions, refs. (37, 38); intracranial recordings: ref. (39)) and with growing evidence for strong integration between structure and lexical meaning in the representations and computations that underlie language processing across fields and approaches, from linguistic theory (e.g., refs. (40–43)), to psycholinguistics (e.g., refs. (44–47)), to computational linguistics (e.g., refs. (48–51)), to cognitive neuroscience (e.g., refs. (10, 39, 52–54)). Furthermore, in line with this strongly lexicalized view of linguistic syntax, although several non-linguistic domains like music, arithmetic, and computer programming exhibit language-like hierarchical structure and have been hypothesized/argued to share combinatorial machinery with language (e.g., refs. (55–58)), growing evidence indicates that functionally distinct brain regions are responsible for structure building in language vs. other domains (e.g., refs. (59–64), see ref. (65) and Fedorenko & Shain, to appear, for reviews).

PDD additionally claim *a distinction between,* on the one hand, *areas that putatively support abstract syntactic processing* (inferior frontal and posterior temporal areas discussed above), *and*, on the other hand, *areas that only support syntactic processing in meaningful* (real-word, not Jabberwocky) *stimuli*. The latter, according to PDD, include anterior temporal areas and the posterior-most parts of the temporal component of the language network (what they refer to as 'TPJ' or temporo-parietal junction—an area that overlaps with our LAngG parcel; **Figure S1-A**). Similarly to PDD, we observe significant interactions between the constituent-length manipulation and stimulus type—with a more pronounced effect of constituent length in real-word than in Jabberwocky stimuli—in the LAntTemp and LAngG language fROIs. However, contra PDD, we observe a) a large and statistically significant constituent-length effect in Jabberwocky stimuli in the LAntTemp language fROI (see also ref. (8); **Figure S1**), b) larger overall responses in the presence of lexical content in the LIFG and LPostTemp language fROIs, and c) a significant constituent length by lexicality interaction in the LIFGorb fROI, along with numerically positive interactions in the LIFG and LPostTemp fROIs (**Figure 2C-D**). Thus, contrary to PDD, our results support similar patterns of response to the critical manipulations across the regions of the language network, rather than PDD's proposed functional subdivision.

The only exception is the LAngG language fROI, which fails to show a significant constituent-length effect for Jabberwocky stimuli, aligning with other studies that have not found sensitivity to structural manipulations therein (e.g., refs. 35, 48) and with studies that have found weaker functional correlations between the LAngG fROI and the rest of the language network (e.g., refs. (25, 34, 67)). The precise role of the LAngG language fROI in linguistic and cognitive processing remains debated, but this region does not appear to be selective for language as it responds more strongly to meaningful pictorial stimuli than to sentences (68, 69).

Going back to the constituent-length effect in real-word stimuli: we asked, what does this effect reflect? We considered the possibility that the empirical predictions of the non-standard sentence processing theory advocated by PDD might be correlated with the predictions of sentence processing theories with wider acceptance and stronger empirical support, thus grounding out PDD's pattern of results in more fundamental explanations of the cognitive mechanisms that underlie language comprehension. We considered several theory-driven measures of sentence processing difficulty (including one, *open nodes*, expressly designed to predict PDD-like build-up effects within constituent strings), and showed that none of them statistically attenuate the constituent-length effect when included as controls, and some of them are actually anti-correlated with constituent length (**Figure 1C**). The constituent-length effect therefore does not align with prominent theories about the influence of syntactic structure on patterns of comprehension difficulty in human sentence processing.

But if the constituent-length effect does not reflect syntactic structure building, what is the right way to interpret this pattern of results? Our proposed answer draws on a prior theoretical distinction between the "proper" and "actual" domains of specialized information processing systems (70, 71), whereby the system's degree of engagement can be modulated by the degree of fit between a given input and the target domain for which the system is adapted. Given the highly combinatory and contextualized nature of natural language, we hypothesize that several words of contiguous context may be necessary in order to identify a stimulus as “proper” to the high-level language system. As a consequence, PDD’s shorter-constituent-length conditions may fail by degrees to fully engage language processing mechanisms in the first place, thereby attenuating overall activation in the language system.

Prior investigations of temporal receptive windows (TRWs, e.g., refs. (33, 72)) support this perspective. A TRW of a brain region (or a voxel, or a neuron) is defined as the length of the preceding context that affects the processing of the current input. Based on the inter-subject correlation approach (73), Blank & Fedorenko (ref. (74); see also ref. (33)) showed that multiword spans of coherent language are needed to maximize synchrony between language network responses across individuals (i.e., to maximize the degree of stimulus-related processing or stimulus ‘tracking’). Relatedly, Fedorenko et al. (39) showed a monotonic increase in activity in electrocorticographic recordings in some language-responsive electrodes over the course of the sentence (see also ref. (75)), with no similar increase shown for strings of unconnected words. These patterns suggest that multiword coherent contexts may be a critical prerequisite for full engagement of the language comprehension system.

Under this view, PDD’s design reveals an effect not because of how syntactic constituents are processed but because of how the language system recognizes inputs as being in-domain. PDD’s design may thus be a parametric variant of contrasts used in other work showing that responses in the language system are diminished by truncation of coherent context (sentences > word lists), removal of lexical content (sentences > Jabberwocky), or both (sentences > non-word lists, or speech > acoustically degraded and thus indecipherable speech) (e.g., refs. (8, 10, 37, 54, 59, 76, 77)). The 2-, 3-, 4-, and 6-word conditions in PDD serve as steps along a continuum of language-likeness between word lists and sentences and correspondingly produce a rise in language network activation.

In summary, our finding of lexicality effects (and/or constituent length by lexicality interactions) in inferior frontal and posterior temporal language regions undermines an influential claim in PDD: that these regions support abstract, content-independent syntactic structure building. Our results are instead consistent with growing evidence that linguistic representations and computations over a range of levels of description (phonological, lexical, syntactic, and semantic) are highly distributed across the language network and are not spatially segregated (10, 52, 54, 74, 78, 79).

We close by noting that despite offering an alternative interpretation of PDD’s findings that does not invoke constituency, our study has no bearing on whether constituency influences human sentence processing in general; we only argue that PDD’s study does not support such an influence. Indeed, abundant evidence for syntactic influences on human sentence processing has accumulated across multiple experimental paradigms (e.g., refs. (66, 80–85)). However, by showing lexicality effects distributed throughout the language system, our results pose a challenge to PDD’s notion of abstract, content-independent syntactic processing centers. Whatever internal functional differentiation the language network may ultimately be shown to exhibit, these results and related evidence of distributed lexical, syntactic, and semantic processing—discussed above—suggest that it is unlikely to be characterized by slicing between levels of linguistic description.

Materials and Methods

This study consists of two replication attempts. Experiment 1 focuses on the real-word conditions from PDD and attempts to replicate the basic constituent-length effect in the language network’s response. Experiment 2 additionally includes Jabberwocky conditions in order to test PDD’s critical theoretical claim: that a subset of the language network implements abstract, content-independent, syntactic processing.

Participants

Forty individuals (age 18-38, 22 females) participated for payment (Experiment 1: n=15; Experiment 2: n=25). All were right-handed—as determined by the Edinburgh Handedness Inventory (86), or self-report—native speakers of English from Cambridge/Boston, MA and the surrounding community. All participants gave informed consent in accordance with the requirements of MIT's Committee on the Use of Humans as Experimental Subjects (COUHES). Each participant completed a language localizer task (8) and a critical task.

Critical Task

The design of both experiments followed PDD but used English materials available at https://osf.io/7pknb/ (the original experiments were carried out in French). In particular, participants were presented with same-length strings (12 words/nonwords), and the internal composition of these strings varied across conditions. The conditions in Experiment 1 were similar to PDD's real-word conditions, except they did not include the 3-word constituent condition. Experiment 2 included three types of experimental manipulation: a) six conditions that were identical to the real-word conditions in PDD: a sequence of twelve unconnected words (i.e., constituents of length 1: c01; here and elsewhere, our condition name abbreviations are similar to those in PDD), six 2-word constituents (c02), four 3-word constituents (c03), three 4-word constituents (c04), two 6-word constituents (c06), and a 12-word sentence (c12); b) three conditions that were a subset of the Jabberwocky conditions from PDD selected to span the range of constituent lengths: a list of twelve unconnected nonwords (jab-c01), three 4-word Jabberwocky constituents (jab-c04), and a 12-word Jabberwocky sentence (jab-c12); and c) two non-constituent conditions (four 3-word non-constituents (nc03) and three 4-word non-constituents (nc04)). We report the results for the non-constituent conditions in **SI Section 9** given that they are not critical for the main question investigated here. Sample stimuli are shown in **Figure 1A**, with the distribution of parts of speech by condition in **Figure 1B**.

Procedure

The procedure was similar for the two experiments and followed PDD: participants saw the stimuli presented one word/nonword at a time in the center of the screen in all caps with no punctuation at the rate of 300 ms per word/nonword (for 3.6 s total trial duration). In Experiment 1, the 150 trials (30 12-word sequences x 5 conditions) were distributed across 5 runs, so that each run contained 6 trials per condition. In addition, each run included 108 s of fixation, for a total run duration of 216 s (3 min 36 s). In Experiment 2, the 330 trials (30 12-word sequences x 11 conditions) were distributed across 10 runs, so that each run contained 3 trials per condition. In addition, each run included 121.2 s of fixation, for a total run duration of 240 s (4 min). In both experiments, the order of conditions and the distribution of fixation periods in each run were determined with the optseq2 algorithm (87).

Linguistic Analyses

The language processing mechanisms that are assumed by PDD commit larger and larger neural assemblies to the representation of a constituent as it is processed word by word, resulting in a hypothesized monotonic increase in language network activation as a function of constituent length (**Figure 1C, PDD**), motivated by evidence of such increases during sequence memory tasks in non-linguistic primates (88–90), and by consistency with a general computational model of distributed associative memory (91). PDD do not attempt to ground their hypothesis in an extensive psycholinguistic literature on the mechanisms of human sentence processing, leaving open the possibility that constituent length effects derive indirectly from more fundamental sentence processing mechanisms. This does not in itself undermine PDD's claim as long as their results are well approximated by measures with external theoretical and empirical support. If—as our results indicate—this is not the case, such an outcome undermines the interpretation of PDD's results as driven by syntactic structure building (see **Discussion** for an alternative account).

To test whether PDD's results are approximated by independently motivated word-by-word measures of processing demand, we considered six alternative measures from the psycholinguistic literature: four of them are derived from memory-based accounts of sentence processing (75, 79, 92), and the other two—from surprisal-based accounts (93–95). The mean value of each predictor by condition is plotted in **Figure 1C**.

Open nodes and *nodes merged* are measures of memory demand hypothesized by (75). They respectively denote costs associated with storing incomplete syntactic constituents and integrating syntactic constituents once they are completed. *Integration cost* and *storage cost* are measures of memory demand hypothesized by the Dependency Locality Theory (DLT) (92). They respectively denote costs associated with keeping incomplete syntactic dependencies in memory (e.g., the awaited verb once the subject is encountered) and retrieving items from memory in order to construct syntactic dependencies to them (i.e. retrieving the subject once the verb is encountered). *5-gram surprisal* and *PCFG surprisal* denote measures of word predictability derived respectively from (i) a computational model that predicts words based on the four preceding words (*5-gram surprisal*), and (ii) a computational model that predicts words based on hypotheses about the sentence's constituent structure (probabilistic context-free grammar or *PCFG surprisal*).

For extended discussion of these predictors and their possible relationship to the constituent length effects reported by PDD, see **SI Section 5**.

Based on theories of working memory demand and surprisal, language processing difficulty (hence, amount of computation/activation) is expected to *increase* with increases in storage/integration costs and surprisal. Because language network activation increases with constituent length, for any variables that underlie PDD's pattern, expected difficulty should also increase with constituent length. This is not the case for the *node closings* predictor, as discussed above, nor for either of the surprisal measures (**Figure 1C**). Anti-correlations with the surprisal measures are not unexpected: words within constituents are expected to be more informative about each other than words that span constituent boundaries, and words with longer contexts are expected to be more predictable on average than words with less context, given that more evidence can accumulate to support the prediction. By contrast, accounting for the influence of function words (larger DLT integration cost at short constituent lengths are driven by the lower proportion of function words in these items, **Figure 1C**), the DLT predictors (and the *open nodes* predictor, as discussed above) show the expected positive association with constituent length. This is again unsurprising: longer constituents permit both longer dependencies (higher integration cost) and more incomplete dependencies at any point in processing (higher storage cost and more open nodes).

Imaging, Functional Localization, and Data Analysis

Imaging, functional localization, and data analysis procedures are described in **SI Sections 1-4**.

Acknowledgments

We acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research, MIT. For technical support during scanning, we thank Steve Shannon and Atsushi Takahashi. We thank Tamar Regev for help with the ROI projection figures, the audience at the

Neurobiology of Language conference in 2020 for helpful discussions of this work, and Rebecca Saxe, Ted Gibson, and Nancy Kanwisher for comments on an earlier version of the manuscript. We also thank Christophe Pallier for sharing the ROI files. E.F. was supported by NIH awards R01-DC016607, R01-DC016950, and U01-NS121471, and by the funds from the McGovern Institute for Brain Research, Brain and Cognitive Sciences Department, and the Simons Center for the Social Brain. Conflicts of Interest: None.

References

1. S. Lewis, C. Phillips, Aligning grammatical theories and language processing models. *J. Psycholinguist. Res.* **44**, 27–46 (2015).
2. C. Pallier, A.-D. Devauchelle, S. Dehaene, Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci.* **108**, 2522–2527 (2011).
3. J. J. Bolhuis, I. Tattersall, N. Chomsky, R. C. Berwick, How could language have evolved? *PLoS Biol* **12**, e1001934 (2014).
4. S. Dehaene, F. Meyniel, C. Wacongne, L. Wang, C. Pallier, The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* **88**, 2–19 (2015).
5. W. T. Fitch, Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Phys. Life Rev.* **11**, 329–364 (2014).
6. I. Bornkessel-Schlesewsky, M. Schlesewsky, S. L. Small, J. P. Rauschecker, Neurobiological roots of language in primate audition: Common computational properties. *Trends Cogn. Sci.* **19**, 142–150 (2015).
7. C. I. Petkov, E. Jarvis, Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* **4**, 12 (2012).
8. E. Fedorenko, P.-J. Hsieh, A. Nieto-Castañón, S. Whitfield-Gabrieli, N. Kanwisher, New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J. Neurophysiol.* **104**, 1177–1194 (2010).
9. E. Fedorenko, A. Nieto-Castañón, N. Kanwisher, Syntactic processing in the human brain: What we know, what we don't know, and a suggestion for how to proceed. *Brain Lang.* **120**, 187–207 (2012).
10. E. Fedorenko, A. Nieto-Castañon, N. Kanwisher, Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia* **50**, 499–513 (2012).
11. J. M. Rodd, M. H. Davis, I. S. Johnsrude, The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb. Cortex* **15**, 1261–1269 (2005).
12. P. Hagoort, L. Hald, M. Bastiaansen, K. M. Petersson, Integration of word meaning and world knowledge in language comprehension. *Science* **304**, 438–441 (2004).
13. C. Rogalsky, G. Hickok, Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cereb. Cortex* **19**, 786–796 (2009).
14. C. Humphries, J. R. Binder, D. A. Medler, E. Liebenthal, Syntactic and semantic

modulation of neural activity during auditory sentence comprehension. *J. Cogn. Neurosci.* **18**, 665–679 (2006).

15. K. Rayner, Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychol. Bull.* **124**, 372–422 (1998).
16. S. Konieczny, On the Difference between Merging Knowledge Bases and Combining them. *KR*, 135–144 (2000).
17. S. Vasishth, R. L. Lewis, Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language (Baltim).* **82**, 767–794 (2006).
18. A. P. Holmes, K. J. Friston, Generalisability, Random Effects & Population Inference. *Neuroimage* **7**, S754 (1998).
19. G. Chen, Z. S. Saad, J. C. Britton, D. S. Pine, R. W. Cox, Linear mixed-effects modeling approach to FMRI group analysis. *Neuroimage* **73**, 176–190 (2013).
20. A. R. Hariri, The neurobiology of individual differences in complex behavioral traits. *Annu. Rev. Neurosci.* **32**, 225–247 (2009).
21. N. Kriegeskorte, W. K. Simmons, P. S. F. Bellgowan, C. I. Baker, Circular analysis in systems neuroscience: The dangers of double dipping. *Nat. Neurosci.* **12**, 535–540 (2009).
22. M. A. Frost, R. Goebel, Measuring structural--functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *Neuroimage* **59**, 1369–1381 (2012).
23. A. M. Tahmasebi, *et al.*, Is the link between anatomical structure and function equally strong at all cognitive levels of processing? *Cereb. cortex* **22**, 1593–1603 (2012).
24. B. Vázquez-Rodríguez, *et al.*, Gradients of structure--function tethering across neocortex. *Proc. Natl. Acad. Sci.* **116**, 21219–21227 (2019).
25. K. Mahowald, E. Fedorenko, Reliable individual-level neural markers of high-level language processing: A necessary precursor for relating neural variability to behavioral and genetic variability. *Neuroimage* **139**, 74–93 (2016).
26. A. Nieto-Castañón, E. Fedorenko, Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage* **63**, 1646–1669 (2012).
27. R. Saxe, M. Brett, N. Kanwisher, Divide and conquer: a defense of functional localizers. *Neuroimage* **30**, 1088–1096 (2006).
28. E. Fedorenko, The early origins and the growing popularity of the individual-subject analytic approach in human neuroscience. *Curr. Opin. Behav. Sci.* **40**, 105–112 (2021).
29. E. Fedorenko, J. Duncan, N. Kanwisher, Language-selective and domain-general regions lie side by side within Broca's area. *Curr. Biol.* **22**, 2059–2062 (2012).
30. S. Shashidhara, F. S. Spronkers, Y. Erez, Individual-subject functional localization increases Univariate activation but not multivariate pattern discriminability in the "multiple-demand" frontoparietal network. *J. Cogn. Neurosci.* **32**, 1348–1368 (2020).
31. R. M. Braga, L. M. DiNicola, H. C. Becker, R. L. Buckner, Situating the left-lateralized

language network in the broader organization of multiple specialized large-scale distributed networks. *J. Neurophysiol.* **124**, 1415–1448 (2020).

32. L. Giglio, M. Ostarek, K. Weber, P. Hagoort, Commonalities and asymmetries in the neurobiological infrastructure for language production and comprehension. *Cereb. Cortex. Adv. online Publ.* (2021).

33. Y. Lerner, C. J. Honey, L. J. Silbert, U. Hasson, Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. *J. Neurosci.* **31**, 2906–2915 (2011).

34. I. Blank, N. Kanwisher, E. Fedorenko, A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *J. Neurophysiol.* **112**, 1105–1118 (2014).

35. E. Fedorenko, S. L. Thompson-Schill, Reworking the language network. *Trends Cogn. Sci.* **18**, 120–126 (2014).

36. M.-M. Mesulam, Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann. Neurol. Off. J. Am. Neurol. Assoc. Child Neurol. Soc.* **28**, 597–613 (1990).

37. M. Bedny, A. Pascual-Leone, D. Dodell-Feder, E. Fedorenko, R. Saxe, Language processing in the occipital cortex of congenitally blind adults. *Proc. Natl. Acad. Sci.* **108**, 4429–4434 (2011).

38. W. Matchin, C. Brodbeck, C. Hammerly, E. Lau, The temporal dynamics of structure and content in sentence comprehension: Evidence from fMRI-constrained MEG. *Hum. Brain Mapp.* **40**, 663–678 (2018).

39. E. Fedorenko, *et al.*, Neural correlate of the construction of sentence meaning. *Proc. Natl. Acad. Sci.* **113**, E6256--E6262 (2016).

40. R. Kaplan, J. Bresnan, “Lexical Functional Grammar: A Formal System for Grammatical Representation” in *The Mental Representation of Grammatical Relations*, J. Bresnan, Ed. (MIT Press, 1982), pp. 173–281.

41. C. Pollard, I. Sag, *Head-driven Phrase Structure Grammar* (University of Chicago Press, 1994).

42. R. Jackendoff, *Semantic structures* (The MIT Press, 1990).

43. A. Goldberg, *Constructions at Work: the nature of generalization in language* (Oxford University Press, 2006).

44. W. Schuler, A. Wheeler, Cognitive Compositional Semantics using Continuation Dependencies in *Third Joint Conference on Lexical and Computational Semantics (*{{SEM}}’14)*, (2014).

45. Y. Kamide, C. Scheepers, G. T. M. Altmann, Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *J. Psycholinguist. Res.* **32**, 37–55 (2003).

46. L. Pylkkänen, B. McElree, “The syntax-semantic interface: On-line composition of sentence meaning” in *Handbook of Psycholinguistics*, M. J. Traxler, M. A. Gernsbacher, Eds. (Elsevier, 2006).

47. M. C. MacDonald, N. J. Pearlmutter, M. S. Seidenberg, The lexical nature of syntactic ambiguity resolution. *Psychol. Rev.* **101**, 676–703 (1994).

48. B.-D. Oh, W. Schuler, Contributions of Propositional Content and Syntactic Category Information in Sentence Processing in *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, (2021), pp. 241–250.

49. T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient Estimation of Word Representations in Vector Space. *CoRR* **abs/1301.3**, 1–12 (2013).

50. C. Dyer, A. Kuncoro, M. Ballesteros, N. A. Smith, Recurrent neural network grammars in *Knight K, Lopez A, Mitchell M, Editors. Human Language Technologies. 2016 Conference of the North American Chapter of the Association for Computational Linguistics; 2016 June 12-17; San Diego (CA, USA).[SI]: Association for Computational Linguistics (ACL)*, (2016).

51. C. Manning, H. Schütze, *Foundations of Statistical Natural Language Processing* (MIT Press, 1999).

52. E. Fedorenko, I. Blank, M. Siegelman, Z. Mineroff, Lack of selectivity for syntax relative to word meanings throughout the language network. *Cognition* **203**, 104348 (2020).

53. I. Blank, Z. Balewski, K. Mahowald, E. Fedorenko, Syntactic processing is distributed across the language system. *Neuroimage* **127**, 307–323 (2016).

54. A. Bautista, S. M. Wilson, Neural responses to grammatically and lexically degraded speech. *Lang. Cogn. Neurosci.* **31**, 567–574 (2016).

55. H. Jeon, Hierarchical processing in the prefrontal cortex in a variety of cognitive domains. *Front. Syst. Neurosci.* **8**, 223 (2014).

56. E. Koechlin, T. Jubault, Broca's Area and the Hierarchical Organization of Human Behavior. *Neuron* **50**, 963–974 (2006).

57. W. T. Fitch, M. D. Martins, Hierarchical processing in music, language, and action: Lashley revisited. *Ann. N. Y. Acad. Sci.* **1316**, 87–104 (2014).

58. F. Lerdahl, R. S. Jackendoff, R. Jackendoff, *A Generative Theory of Tonal Music* (MIT Press, 1983).

59. E. Fedorenko, M. K. Behr, N. Kanwisher, Functional specificity for high-level linguistic processing in the human brain. *Proc. Natl. Acad. Sci.* (2011).

60. A. A. Ivanova, *et al.*, Comprehension of computer code relies primarily on domain-general executive brain regions. *Elife* **9**, e58906 (2020).

61. M. M. Monti, L. M. Parsons, D. N. Osherson, Thought beyond language: Neural dissociation of algebra and natural language. *Psychol. Sci.* **23**, 914–922 (2012).

62. M. Amalric, S. Dehaene, A distinct cortical network for mathematical knowledge in the human brain. *Neuroimage* **189**, 19–31 (2019).

63. X. Chen, *et al.*, The human language system does not support music processing. *bioRxiv* (2021).

64. Y.-F. Liu, J. Kim, C. Wilson, M. Bedny, Computer code comprehension shares neural

resources with formal logical inference in the fronto-parietal network. *Elife* **9**, e59340 (2020).

65. E. Fedorenko, I. Blank, Broca's Area Is Not a Natural Kind. *Trends Cogn. Sci.* (2020).

66. C. Shain, I. Blank, M. van Schijndel, W. Schuler, E. Fedorenko, fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia* **138**, 107307 (2020).

67. L. R. Chai, M. G. Mattar, I. A. Blank, E. Fedorenko, D. S. Bassett, Functional network dynamics of the language system. *Cereb. Cortex* **26**, 4148–4159 (2016).

68. E. Amit, C. Hoeflin, N. Hamzah, E. Fedorenko, An asymmetrical relationship between verbal and visual thinking: Converging evidence from behavior and fMRI. *Neuroimage* **152**, 619–627 (2017).

69. A. A. Ivanova, *et al.*, The language network is recruited but not required for nonverbal event semantics. *Neurobiol. Lang.* **2**, 176–201 (2021).

70. D. Sperber, The modularity of thought and the epidemiology of representations. *Mapp. mind Domain Specif. Cogn. Cult.*, 39–67 (1994).

71. H. C. Barrett, R. Kurzban, Modularity in cognition: framing the debate. *Psychol. Rev.* **113**, 628 (2006).

72. U. Hasson, E. Yang, I. Vallines, D. J. Heeger, N. Rubin, A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* **28**, 2539–2550 (2008).

73. U. Hasson, R. Malach, D. J. Heeger, Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* **14**, 40–48 (2010).

74. I. Blank, E. Fedorenko, No evidence for differences among language regions in their temporal receptive windows. *Neuroimage*, 116925 (2020).

75. M. J. Nelson, *et al.*, Neurophysiological dynamics of phrase-structure building during sentence processing. *Proc. Natl. Acad. Sci.* **114**, E3669--E3678 (2017).

76. B. Röder, O. Stock, H. Neville, S. Bien, F. Rösler, Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: A functional magnetic resonance imaging study. *Neuroimage* **15**, 1003–1014 (2002).

77. T. M. Snijders, *et al.*, Retrieval and unification of syntactic structure in sentence comprehension: an fMRI study using word-category ambiguity. *Cereb. cortex* **19**, 1493–1503 (2009).

78. T. I. Regev, *et al.*, High-level language brain regions are sensitive to sub-lexical regularities. *bioRxiv* (2021).

79. C. Shain, I. A. Blank, E. Fedorenko, E. Gibson, W. Schuler, Robust effects of working memory demand during naturalistic language comprehension in language-selective cortex. *bioRxiv* (2021) https:/doi.org/10.1101/2021.09.18.460917.

80. T. G. Bever, J. R. Lackner, R. Kirk, The underlying structures of sentences are the primary units of immediate speech processing. *Percept. Psychophys.* **5**, 225–234 (1969).

81. D. Caplan, C. Baker, F. Dehaut, Syntactic determinants of sentence comprehension in

aphasia. *Cognition* **21**, 117–175 (1985).

82. B. McElree, The locus of lexical preference effects in sentence comprehension: A time-course analysis. *J. Mem. Lang.* **32**, 536–571 (1993).

83. J. A. Van Dyke, B. McElree, Retrieval interference in sentence comprehension. *J. Mem. Lang.* **55**, 157–166 (2006).

84. D. J. Grodner, E. Gibson, Consequences of the serial nature of linguistic input. *Cogn. Sci.* **29**, 261–291 (2005).

85. V. Fossum, R. Levy, Sequential vs. Hierarchical Syntactic Models of Human Incremental Sentence Processing in *Proceedings of {{CMCL}} 2012*, (Association for Computational Linguistics, 2012).

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

87. A. M. Dale, B. Fischl, M. I. Sereno, Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).

88. P. Barone, J.-P. Joseph, Prefrontal cortex and spatial sequencing in macaque monkey. *Exp. brain Res.* **78**, 447–464 (1989).

89. Y. Ninokura, H. Mushiake, J. Tanji, Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. *J. Neurophysiol.* **89**, 2868–2873 (2003).

90. Y. Ninokura, H. Mushiake, J. Tanji, Integration of temporal order and object information in the monkey lateral prefrontal cortex. *J. Neurophysiol.* **91**, 555–560 (2004).

91. P. Smolensky, G. Legendre, *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar* (MIT Press, 2006).

92. E. Gibson, “The Dependency Locality Theory: A distance-based theory of linguistic complexity” in *Image, Language, Brain*, A. Marantz, Y. Miyashita, W. O’Neil, Eds. (MIT Press, 2000), pp. 95–106.

93. J. Hale, A Probabilistic Earley Parser as a Psycholinguistic Model in *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*, (2001), pp. 159–166.

94. R. Levy, Expectation-based syntactic comprehension. *Cognition* **106**, 1126–1177 (2008).

95. M. van Schijndel, A. Exley, W. Schuler, A model of language processing as hierarchic sequential prediction. *Top. Cogn. Sci.* **5**, 522–540 (2013).

Figures and Tables

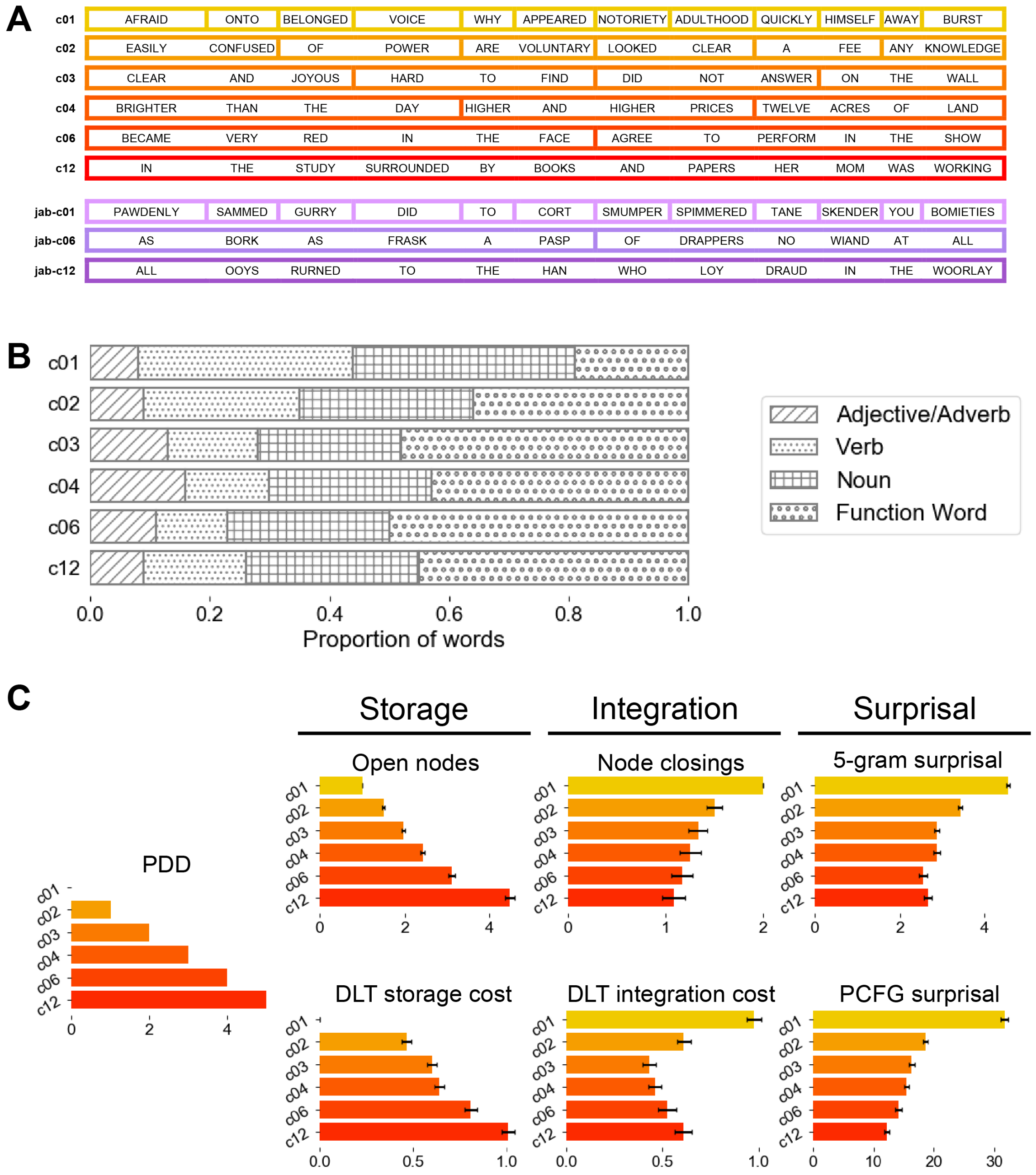

Figure 1. (**A**) Examples of stimuli across length conditions (from 1-word constituents, c01, to 12-word constituents, c12), with real-word constituent conditions shown in warm colors and Jabberwocky constituent conditions shown in cool colors. (**B**) Proportion of parts of speech (PoS) by constituent length for real-word conditions. At length 1, nouns and verbs are overrepresented and function words are underrepresented. This is because function words easily license multiword constituents, violating the 1-word constraint. At lengths 3+, the distribution of categories is relatively stable. (**C**) Mean value of linguistic features (memory- and surprisal-based) by constituent length for real-word conditions, compared to PDD-hypothesized monotonic increase (left). Error bars show standard errors of the mean across items.

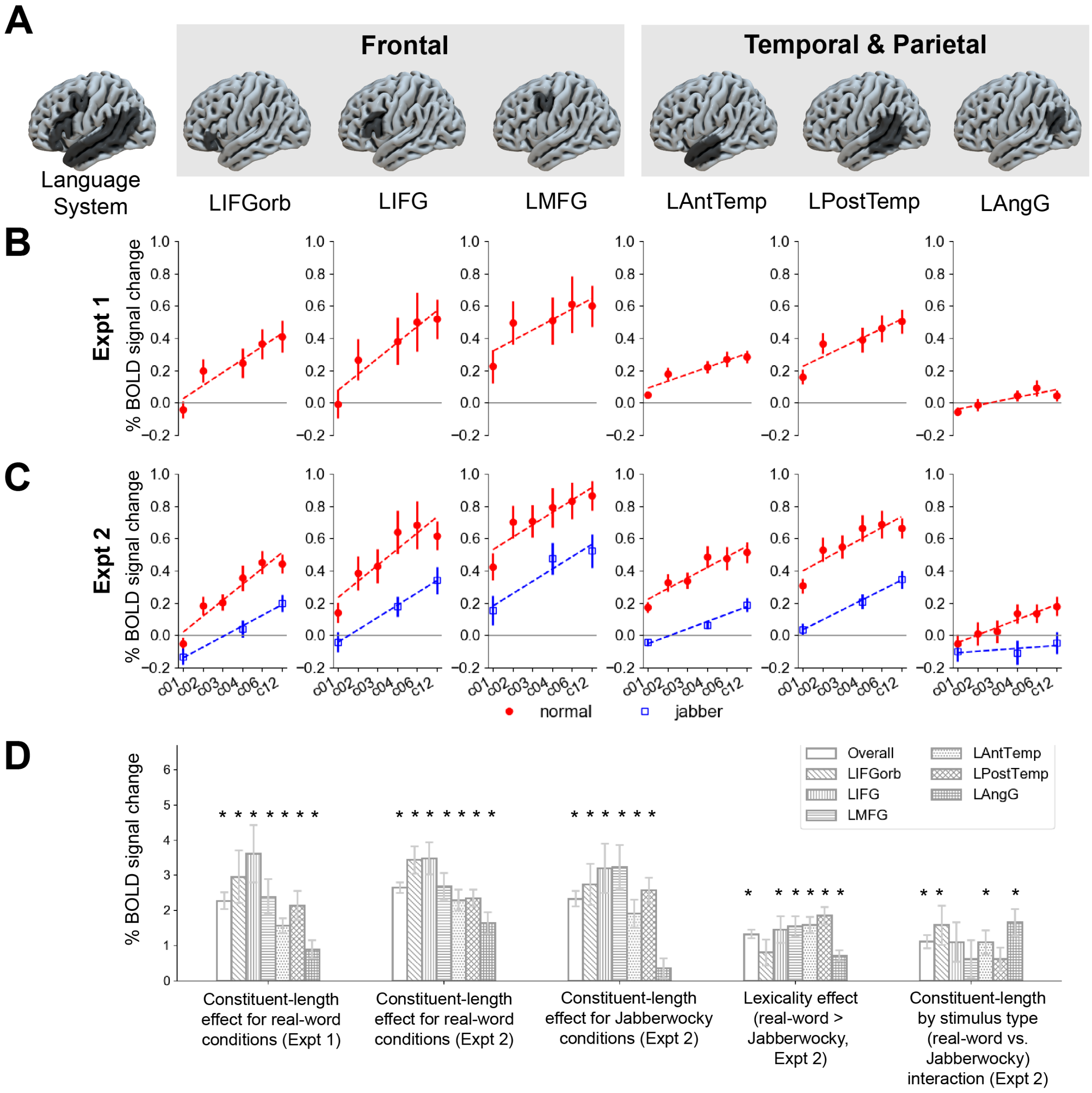

Figure 2. A. Group masks bounding the six left-hemisphere regions of the language network. The top 10% of language-selective voxels are selected within each mask in each participant. **B.** Estimated response to each condition of the real-words conditions in Expt 1 (which did not include Jabberwocky conditions). Responses in all regions increase with constituent length. **C.** Estimated response to each condition of the real-words conditions (replicating Expt 1) and the Jabberwocky conditions in Expt 2. Responses in all regions increase with constituent length in both real-word and Jabberwocky conditions. **D.** Key contrasts by language network fROI (left-to-right): constituent-length effect for real-word conditions in Expt 1; constituent-length effect for real-word conditions in Expt 2; constituent-length effect for Jabberwocky conditions in Expt 2; overall lexicality effect (increase in response for real-word over Jabberwocky conditions in Expt 2, averaging over length); increase in constituent-length effect in real-word conditions over Jabberwocky in Expt 2. Starred bars indicate statistically significant effects by likelihood ratio test. (Note that effect sizes in **D** are not always identical to their corresponding slopes in **B** and **C** because some contrasts use a subset of available length conditions for valid comparison.) Error bars show standard error of the mean over participants.

	sensitivity to lexical content (inconsistent with abstract syntactic processsing)		**sensitivity to structure in Jabberwocky**	
	PDD	SKLASMF	PDD	SKLASMF
inferior frontal	−	+	+	+
anterior temporal	+	+	−	+
posterior temporal	−	+	+	+
AngG/TPJ	+	+	−	−

Table 1. Summary of key similarities and differences between PDD's findings and those of our study (SKLASMF). PDD reported (a) one set of regions (inferior frontal and posterior temporal) that were sensitive to structure (constituent length) in real-word stimuli and equally sensitive to structure in Jabberwocky stimuli (supporting abstract syntactic processing in these regions), and (b) another set of regions (anterior temporal and TPJ) that were sensitive to lexical content and insensitive to structure in Jabberwocky stimuli. Our study does not reproduce most of PDD's reported insensitivities (red minus signs), instead finding sensitivity to both lexical content and syntactic structure in Jabberwocky stimuli throughout the regions of the functional language network.