Neural correlate of the construction of sentence meaning

Evelina Fedorenko[a,b,1], Terri L. Scott[c], Peter Brunner[d,e], William G. Coon[d,f], Brianna Pritchett[g], Gerwin Schalk[d,e,f], and Nancy Kanwisher[g,1]

[a]Department of Psychiatry, Harvard Medical School, Boston, MA 02115; [b]Department of Psychiatry, Massachusetts General Hospital, Boston, MA 02114; [c]Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA 02215; [d]National Center for Adaptive Neurotechnologies, Wadsworth Center, New York State Department of Health, Albany, NY 12208; [e]Department of Neurology, Albany Medical College, Albany, NY 12208; [f]Department of Biomedical Sciences, State University of New York at Albany, Albany, NY 12222; and [g]Brain & Cognitive Sciences Department, McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Nancy Kanwisher, August 4, 2016 (sent for review February 17, 2016; reviewed by Daphne Bavelier and Ziv M. Williams)

The neural processes that underlie your ability to read and understand this sentence are unknown. Sentence comprehension occurs very rapidly, and can only be understood at a mechanistic level by discovering the precise sequence of underlying computational and neural events. However, we have no continuous and online neural measure of sentence processing with high spatial and temporal resolution. Here we report just such a measure: intracranial recordings from the surface of the human brain show that neural activity, indexed by γ-power, increases monotonically over the course of a sentence as people read it. This steady increase in activity is absent when people read and remember nonword-lists, despite the higher cognitive demand entailed, ruling out accounts in terms of generic attention, working memory, and cognitive load. Response increases are lower for sentence structure without meaning ("Jabberwocky" sentences) and word meaning without sentence structure (word-lists), showing that this effect is not explained by responses to syntax or word meaning alone. Instead, the full effect is found only for sentences, implicating compositional processes of sentence understanding, a striking and unique feature of human language not shared with animal communication systems. This work opens up new avenues for investigating the sequence of neural events that underlie the construction of linguistic meaning.

language | ECoG | compositionality | syntax | semantics

How does a sequence of sounds emerging from one person's mouth create a complex meaning in another person's mind? Although we have long known where language is processed in the brain (1–3), we still know almost nothing about how neural circuits extract and represent the meaning of a sentence. A powerful method for addressing this question is intracranial recording of neural activity directly from the cortical surface in neurosurgery patients (i.e., electrocorticography or ECoG) (4, 5). Although opportunities for ECoG data collection are rare, determined by clinical—not scientific—priorities, they nonetheless offer an unparalleled combination of spatial and temporal resolution, and further provide direct measures of actual neural activity, rather than indirect measures via blood flow (as in PET, fMRI, and near infrared spectroscopy/optical imaging). ECoG data are particularly valuable for the study of uniquely human functions like language, where animal models are inadequate. Here we used ECoG to identify the neural events that occur online as the meaning of a sentence is extracted and represented.

Prior intracranial recording studies of language have largely focused on speech perception and production (e.g., refs. 6–11) and word-level processes (e.g., refs. 12–26). However, the most distinctive feature of human language is its compositionality: the ability to create and understand complex meanings from novel combinations of words structured into phrases and sentences (27). As a first step toward understanding the neural basis of sentence comprehension, we recorded intracranial responses while participants read sentences and three kinds of control stimuli. In each trial, a string of eight items (words or nonwords) were presented sequentially, enabling us to separately measure the neural response to each item individually; to encourage attention to all stimuli (whether meaningful or not), participants decided at the end of each trial whether a probe word/nonword appeared in the preceding string.

Our four stimulus conditions (materials adapted from ref. 28; examples are shown in the "Sample materials" table below) enabled us to orthogonally vary the presence of word meaning and sentence structure. Word-lists (W) included the same words as the sentences (S) (scrambled across sentences), but lacked sentence structure. "Jabberwocky" (J) sentences were grammatical but largely meaningless, as they contained no real content words. Nonword-lists (N) contained neither sentence-level structure nor word meaning.

Sample materials

Condition	Example
Sentences	STEVE WAS LATE TO SCHOOL BECAUSE HE OVERSLEPT [probe: SCHOOL] THE RED BALLOON ROSE UP INTO THE CLOUDS [probe: WENT]
Word-lists	RAIN THE WORK BEHIND REACHED GREW KIDS OPENED [probe: GREW] STOOD THE TIED CANDLE INTO SHED THE QUICKLY [probe: WALLET]
Jabberwocky	THE GAR WAS SWARBING THE MUME FROM ATAR [probe: ATAR] TOMAL HOTHED THE BLESPY NULO DURING THE VAYLANT [probe: FLORKY]
Nonword-lists	PHREZ CRE EKED PICUSE EMTO PECH CRE ZEIGELY [probe: PHREZ] PIV WUBA WOS PAFFING DEBON TRIENED LE KIF [probe: LOME]

Significance

How do circuits of neurons in your brain extract and hold the meaning of a sentence? To start to address this unanswered question, we measured neural activity from the surface of the human brain in patients being mapped out before neurosurgery, as they read sentences. In many electrodes, neural activity increased steadily over the course of the sentence, but the same was not found when participants read lists of words or pronounceable nonwords, or grammatical nonword strings ("Jabberwocky"). This build-up of neural activity appears to reflect neither word meaning nor syntax alone, but the representation of complex meanings.

Author contributions: E.F., G.S., and N.K. designed research; E.F., T.L.S., P.B., W.G.C., and G.S. performed research; E.F., T.L.S., P.B., B.P., G.S., and N.K. analyzed data; and E.F., T.L.S., G.S., and N.K. wrote the paper.

Reviewers: D.B., University of Geneva; and Z.M.W., Harvard Medical School.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: evelina.fedorenko@mgh.harvard.edu or ngk@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1612132113/-/DCSupplemental.

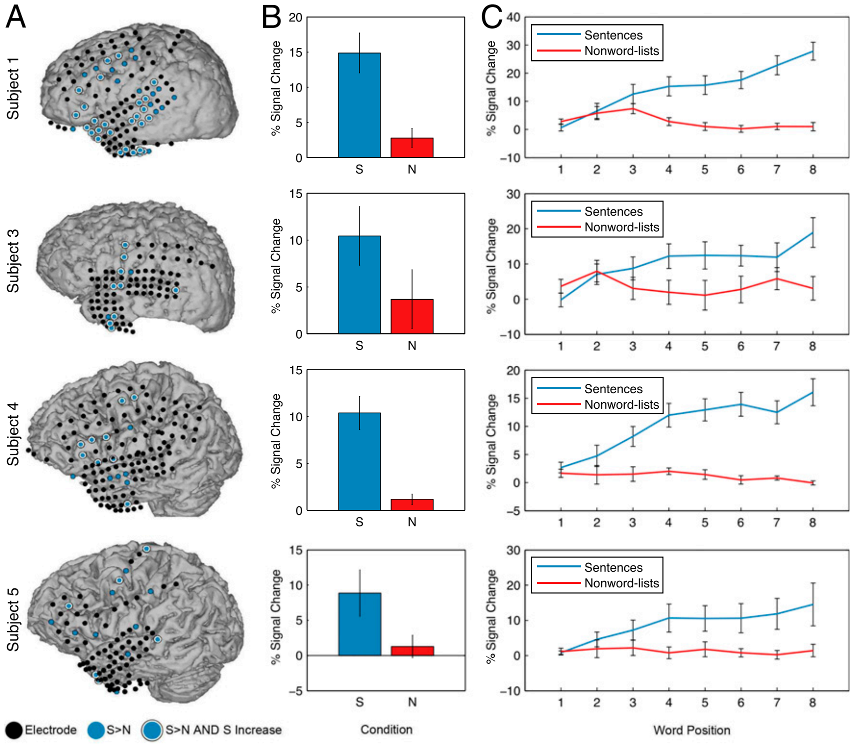

Fig. 1. (*A*) Cortical models of individual subjects showing all electrodes (black) (see Table 1 for numbers of electrodes), electrodes that show a significant sentences > nonword-lists (S > N) effect in odd-numbered runs (blue), and electrodes that show both a significant S > N effect and a monotonic increase across word positions in the sentence condition in odd-numbered runs, [i.e., our electrodes of interest, EOIs (blue, circled in white)]. (*B*) The γ-magnitude for sentences and nonword-lists averaged across word positions estimated in even-numbered runs (i.e., data independent from the data used to select the EOIs). (*C*) The γ-magnitude for sentences and nonword-lists in each of eight word positions in even-numbered runs (see *SI Appendix, Part F* for additional figures showing data not averaged within each word position). Error bars indicate SEMs over EOIs in both *B* and *C*.

Few prior studies have recorded intracranial responses during sentence comprehension. Noninvasive methods with high temporal resolution [event-related brain potential (ERPs) and magnetoencephalography] have revealed neural responses to specific types of events during sentence comprehension, such as syntactic or semantic violations (e.g., refs.29–36), but do not provide a measure of ongoing neural activity from focal cortical regions. Numerous prior studies with fMRI have identified cortical regions that respond strongly (e.g., refs. 28, 37–41) and selectively (42–44) during sentence comprehension (whether presented visually or auditorily) (28, 45), but fMRI lacks the temporal resolution to reveal the responses to individual words as a sentence is presented. Hence, little is known about how neural activity in specific cortical regions unfolds over the course of a sentence as the meaning of that sentence is extracted and represented.

Given the necessarily exploratory nature of this study, we took stringent measures to guard against the dangers of statistical nonindependence and hidden degrees-of-freedom (46): all hypotheses, data analysis choices, and selection of specific electrodes were made based on analyses of only half the data (odd-numbered runs), before the other half of the data (even-numbered runs) were ever inspected.

Six epilepsy patients with subdurally implanted electrodes placed over left-hemisphere frontal, temporal, and parietal cortices were tested. We measured the time-course at each electrode of broadband γ-activity of the ECoG signal, which is closely related to spiking activity of neuronal populations directly underneath each recording electrode (47–50).

Results

We first asked whether any electrodes produced a higher γ-response overall during the reading of sentences than nonword strings, as has been reported in numerous previous studies with fMRI (e.g., refs. 28, 51, 52). As expected, numerous electrodes in each subject, distributed across temporal and frontal regions,

Table 1. Numbers of total, analyzed, and language-responsive electrodes, as well as EOIs

Subject no.	Total electrodes	Analyzed electrodes	S > N Electrodes	S > N and S Increasing (i.e.,EOIs)
S1	120	117	45 (0.38*)	27 (0.23†, 0.60‡)
S3	112	84	11 (0.13)	9 (0.11, 0.82)
S4	134	124	15 (0.12)	9 (0.07, 0.60)
S5	98	87	15 (0.17)	6 (0.06, 0.33)

*Proportion of S > N electrodes relative to analyzed electrodes.
†Proportion of S > N and S increasing electrodes relative to analyzed electrodes.
‡Proportion of S > N and S increasing electrodes relative to S > N electrodes.

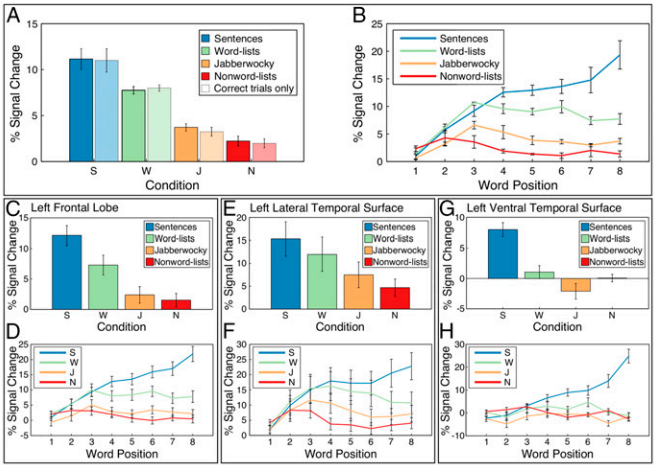

Fig. 2. (*A*) The γ-magnitude for all conditions (S, sentences; W, word-lists; J, Jabberwocky sentences; N, nonword-lists) averaged across word positions and subjects estimated in even-numbered runs. Darker bars show responses across all trials, lighter bars show responses across the subset of trials for which the memory-probe task was answered correctly. (*B*) The γ-magnitude for all conditions averaged across subjects in each of eight word positions in even-numbered runs. (*C*–*H*) The γ-magnitude for all conditions averaged across word positions (*C*, *E*, *G*) and in each of eight word positions (*D*, *F*, *H*) in even-numbered runs in three groups of EOIs: left frontal (*C* and *D*), left lateral temporal (*E* and *F*), and left ventral temporal (*G* and *H*). All four subjects had some EOIs in left frontal (total: 21 EOIs) and left lateral temporal (total: 20 EOIs) cortex, and two of the four had EOIs in left ventral temporal (total: 9 EOIs) cortex (see *SI Appendix, Part G* for numbers of EOIs in each group for each subject). Error bars indicate SEMs over subjects in all panels.

responded significantly more strongly to sentences than nonword strings (Fig. 1*A* and Table 1), when odd-run data only were analyzed (see *SI Appendix, Part I* for evidence that some electrodes show the opposite pattern, plausibly reflecting sensitivity to task difficulty).

The most striking and novel finding from the analyses of odd-run data was that over half of the language-responsive electrodes further showed a monotonic increase of γ-power over the eight words in the sentence condition, while failing to increase for the control nonword-list condition. This monotonic increase appears to reflect a previously unreported marker of the neural construction of sentence meaning.

To rigorously test the significance of this finding, for each subject we selected electrodes of interest (EOIs) (Fig. 1*A* and Table 1) that showed (in odd-run data): (*i*) a significantly greater mean response to sentences than nonword-lists averaging across the eight word positions, and (*ii*) a monotonic increase over the eight positions in the sentence condition (*Materials and Methods*). Fifty-one such EOIs were identified across four subjects, distributed across the frontal and temporal cortex. These four subjects were used in all of the analyses reported in the text (Figs. 1–4, except Fig. 2 *G* and *H*). The remaining two subjects—with only two EOIs each—were excluded from the main analyses, but their results were qualitatively and quantitatively similar (*SI Appendix, Part D*). Response magnitudes of each EOI in each subject were then quantified from even-run data for each condition.

Indeed, the even-run data replicated the higher response to sentences than nonword-lists ($P < 0.005$) (Fig. 1*B* and Table 2). Critically, each subject also replicated the monotonic increase in γ-power for sentences (correlation between word position and response, $P < 0.05$), but not nonword-lists (Fig. 1*C* and *SI Appendix, Part A*; see also *SI Appendix, Part B* for similar results when using even-numbered runs for EOI selection). Reliable differences between the two conditions emerged between the second and fourth word position within the sequence (*SI Appendix, Part A*), suggesting that the ability to combine two or three words into a coherent representation was necessary for the sentence time-course to diverge significantly from that of the control, nonword-lists condition.

Sentences differ from nonword-lists in the presence of both word meaning and syntax. Is the observed response increase primarily driven by one of these factors? To find out, we measured responses to the two remaining conditions: word-lists (lexical meaning with little syntax) and Jabberwocky sentences (syntax with little meaning) (*Sample Materials*). Averaging across word positions, responses are highest for sentences, weaker for word-lists and Jabberwocky, and weakest for nonword-lists (Fig. 2*A* and Table 2), a pattern similar to the one observed previously in fMRI (28, 51). It is worth noting that unlike in fMRI, where the responses to word-lists and Jabberwocky are similar in magnitude (28), the ECoG response to word-lists is generally higher than the response to Jabberwocky (Fig. 2*A*), and significantly so in many EOIs. In fact, across EOIs, each of the four subjects shows a significantly higher response to word-lists than Jabberwocky ($P < 0.05$) (Table 2). No EOI shows a

Table 2. Results from two-tailed paired samples *t* tests on the differences in PSC (with respect to baseline fixation) between pairs of experimental conditions

Subject	S–N	S–W	S–J	W–N	J–N	W–J
S1	12.11 ± 1.78 $t(26) = 6.80$ $P < 10^{-6}$	6.19 ± 0.58 $t(26) = 10.67$ $P < 10^{-10}$	11.42 ± 1.02 $t(26) = 11.21$ $P < 10^{-10}$	5.92 ± 1.92 $t(26) = 3.08$ $P < 10^{-2}$	0.69 ± 1.21 $t(26) = 0.57$ $P = 0.58$	5.23 ± 0.92 $t(26) = 5.65$ $P < 10^{-5}$
S3	6.76 ± 1.02 $t(8) = 6.58$ $P < 10^{-3}$	3.72 ± 0.89 $t(8) = 4.18$ $P < 10^{-2}$	5.42 ± 0.81 $t(8) = 6.71$ $P < 10^{-3}$	3.04 ± 0.65 $t(8) = 4.67$ $P < 10^{-2}$	1.33 ± 0.55 $t(8) = 2.43$ $P = 0.04$	1.71 ± 0.23 $t(8) = 7.38$ $P < 10^{-4}$
S4	9.21 ± 1.38 $t(8) = 6.66$ $P < 10^{-3}$	3.17 ± 0.83 $t(8) = 3.80$ $P < 10^{-2}$	7.59 ± 1.11 $t(8) = 6.83$ $P < 10^{-3}$	6.04 ± 1.47 $t(8) = 4.12$ $P < 10^{-2}$	1.62 ± 0.59 $t(8) = 2.75$ $P = 0.03$	4.42 ± 1.03 $t(8) = 4.31$ $P < 10^{-2}$
S5	7.58 ± 3.51 $t(5) = 2.16$ $P = 0.08$	0.51 ± 1.31 $t(5) = 0.39$ $P = 0.72$	5.24 ± 2.29 $t(5) = 2.29$ $P = 0.07$	7.08 ± 2.73 $t(5) = 2.59$ $P = 0.05$	2.34 ± 1.30 $t(5) = 1.80$ $P = 0.13$	4.73 ± 1.60 $t(5) = 2.96$ $P = 0.03$
Across subjects	8.92 ± 1.18 $t(3) = 7.56$ $P < 10^{-2}$	3.40 ± 1.16 $t(3) = 2.91$ $P = 0.06$	7.42 ± 1.44 $t(3) = 5.17$ $P = 0.01$	5.52 ± 0.87 $t(3) = 6.38$ $P < 10^{-2}$	1.50 ± 0.34 $t(3) = 4.37$ $P = 0.02$	4.02 ± 0.79 $t(3) = 5.10$ $P = 0.01$

Data are taken from even-numbered runs only and averaged across all word positions and EOIs. The first row of each cell denotes the average PSC difference with SEMs across EOIs. Significance levels are shaded with white: $P < 0.01$, midgray: $0.01 \leq P < 0.05$, and dark-gray: $P \geq 0.05$.

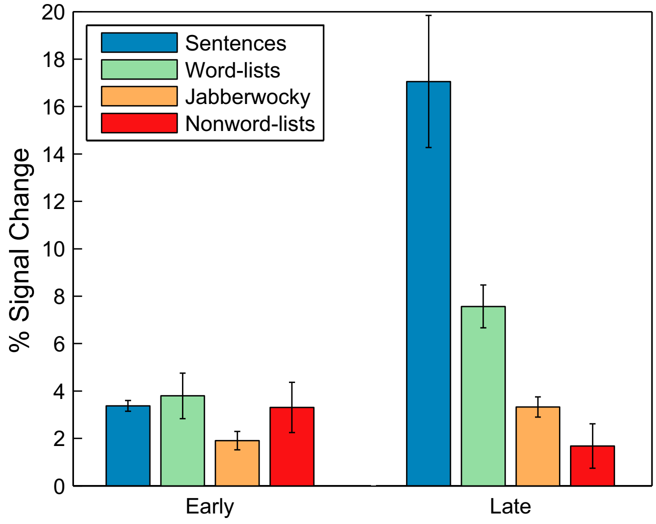

Fig. 3. High γ-magnitude during the early (first two words) and late (last two words) sentence positions estimated across EOIs in even-numbered runs for all conditions. Error bars indicate SEMs over subjects.

significant Jabberwocky > word-lists effect. This asymmetry is consistent with—though not necessarily directly related to—the more robust pattern information for lexical information than syntactic information revealed by multivoxel pattern analyses in fMRI (53).

Critically, the time-courses reveal that neither syntax nor lexical meanings on their own are sufficient to account for the full response increase in the sentence condition (Fig. 2*B*). This conclusion was supported by several two-way repeated-measures ANOVAs with two within-subjects factors: (*i*) condition and (*ii*) position (early = average of the first two words vs. late = average of the last two words). This approach allowed us to test whether the build-up effect in the sentence condition is significantly different from the patterns observed in the other conditions. Including all four conditions, we find a main effect of condition [$F(3, 9) = 11.99$, $P = 0.005$], a main effect of position [$F(1, 3) = 70.08$, $P < 0.005$], and a highly reliable condition by position interaction [$F(3, 9) = 18.18$, $P < 0.001$]. Further, 2 × 2 ANOVAs comparing sentences against each of the other conditions revealed reliable interactions for sentences vs. nonword-lists [$F(1, 3) = 22.12$, $P = 0.0182$] and sentences vs. Jabberwocky [$F(1, 3) = 23.23$, $P = 0.017$], and a marginal interaction for sentences vs. word-lists [$F(1, 3) = 9.551$, $P = 0.0537$] (Fig. 3).

Although both the Jabberwocky and word-list conditions show an increase over the first three word positions, the response subsequently decreases, and correlations between word position and γ were therefore not significant for either condition (Fig. 2*B* and *SI Appendix, Part A*). The initial increase plausibly reflects participants' attempts to interpret the linguistic signal, a natural tendency given our general experience with language, where words combine to create complex meanings. However, upon realizing that no coherent interpretation is possible, these attempts are likely abandoned. Thus, the continuous response increase in the sentence condition apparently reflects neither syntactic processing nor word-level meaning alone, but the compositional construction of sentence meaning. Consistent with this high-level interpretation of the monotonic increase for sentences but not nonwords, the effect is also seen when the stimuli are presented auditorily rather than visually (*SI Appendix, Part H*).

Examining the responses of individual EOIs across subjects suggests that conditions pattern similarly regardless of the EOI's anatomical location. To examine potential anatomical differences more systematically, we split our EOIs into three groups: frontal lobe EOIs, EOIs located on the lateral temporal surface, and EOIs located on the ventral temporal surface (see *SI Appendix, Part G* for numbers of EOIs in each group). As shown in Fig. 2 *C–F*, the data patterns look remarkably similar for the frontal and lateral temporal EOIs. The left ventral temporal electrodes (Fig. 2 *G* and *H*) appear to show greater specificity, with strong build-up for sentences and none at all for the three other conditions. Although intriguing, only two of the four subjects contributed EOIs to this analysis, so it is not clear yet whether this apparent difference in functional response across anatomical locations reflects a general pattern. The distributed nature of the build-up effect contrasts with proposals in the neuroimaging literature that one particular region within the language network houses the core combinatorial/syntactic apparatus: for example, parts of the inferior frontal gyrus (e.g., refs. 54–56) or parts of the anterior temporal cortex (e.g., refs. 57–60). However, it accords with fMRI work that observes sensitivity to lexical and syntactic processing throughout the language network (e.g., refs. 28, 51, 53, 61, 62), and evidence that language regions form a highly integrated functional system (e.g., ref. 63).

Discussion

Our most striking finding is that many language-responsive electrodes show a monotonic increase in γ-power over the course of the sentence as it is read or heard, and that this increase cannot be explained by the presence of either word meaning or sentence structure alone. What mental processes might underlie this increase in γ response? We can rule out several potential explanations linked to general cognitive factors. First, the build-up effect cannot be explained by general attention or arousal that is higher in the sentence condition than in the other three conditions. Performance on the memory-probe task provides a useful proxy for attention/arousal, and participants generally performed well on the memory-probe task across conditions (*SI Appendix, Part C*). Although accuracies were numerically higher in the sentence and word-list conditions than in the Jabberwocky and nonword-list conditions, performance was quite good even in the least-accurate (nonword-list) condition [>70%, not including subject 5 (S5), who misunderstood the instructions] (*SI Appendix, Part C*). These data accord with the subjective impression of performing this task: the nonword-list condition is the hardest, the Jabberwocky and word-list conditions intermediate, and the sentence condition is the easiest. This pattern goes in the opposite direction of that predicted by an attention/arousal/difficulty account. (It is worth noting that a number of electrodes do show a pattern of response indicative of sensitivity to difficulty/effort, with greater responses to nonword-list and Jabberwocky conditions than the sentence condition (*SI Appendix, Part I*), in line with prior fMRI findings (64).

Second, the build-up effect cannot be the result of an overall better performance on the memory-probe task in the sentence condition. One could, in principle, imagine that trials where the memory-probe task is answered correctly are the ones that show an increase over the course of the sentence. Because there are more correctly answered trials in the sentence condition, the build-up effect may emerge in the sentence condition, but not other conditions. This possibility is already unlikely, given the pattern of accuracies across conditions: for example, the word-list condition does not show a significant build-up despite the fact that the accuracies for that condition are similar to those of the sentence condition. However, to test this possibility directly, we compared the time-courses of high γ-responses for correctly vs. incorrectly answered trials in each of the three conditions that failed to show a build-up effect (we could not perform this comparison for the sentence condition given that there were hardly any incorrectly answered trials). As Fig. 4 shows, the increase does not appear to depend on whether the trial was answered correctly: in the conditions other than the sentence condition, neither the correctly nor the incorrectly answered

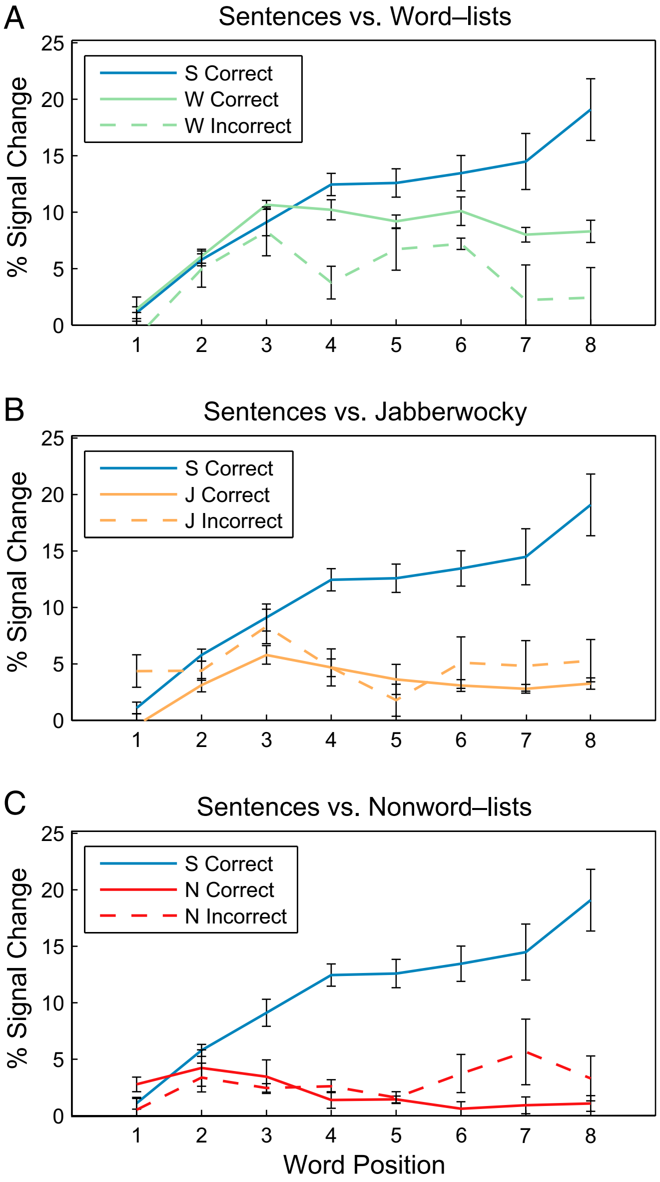

Fig. 4. Comparison of percent signal change (PSC) of correctly and incorrectly answered trials in the word-lists (*A*), Jabberwocky (*B*), and nonword-lists (*C*) conditions, averaged across subjects. For comparison, we include the average of the correctly answered sentence trials (there were too few subjects with incorrectly answered sentence trials to compute averages). Error bars indicate SEMs over subjects.

trials show an increase across positions. Fig. 2*A* further shows the overall magnitude for the correctly answered trials across the four conditions (lighter bars): the pattern of results looks almost identical to the analysis that includes all trials (darker bars).

Third, the build-up effect cannot be caused by the anticipation of the end of the trial because across conditions all trials were eight words/nonwords-long, thus the patterns should be similar across conditions. Furthermore, according to this hypothesis, one would predict a drop on the last word of the sentences—where the end of the trial is reached—which we do not observe (instead, the response keeps increasing). Similarly, the effect cannot be caused by the preparation to perform the memory-probe task, because the task is the same across conditions.

What about hypotheses that more specifically invoke language processing? Words that occur later in the sentence are more predictable (65–67), and hence easier to process. However, it is unclear why neural activity would increase for words that are easier to process, especially given that the N400 ERP component decreases for words in later sentence positions (67), as do reading times (68, 69). Another possibility is that the response increase reflects strengthening predictions about upcoming material. Evidence against this possibility is that the signal did not decrease substantially at points where a complete clause is formed, and thus no further material is necessary/predicted, or at the following word (*SI Appendix*, *Part E*). Thus, the most likely explanation of the response increase is that it reflects the increasing complexity of the evolving representation of the meaning of the sentence. In particular, because items must be held in memory for all conditions, and the steady increase in γ is greatest for sentences, that increase must reflect the representation of structured sentence meaning over and above the representation of individual lexical items.

It may be useful to link these results to classic spreading-activation theories of lexico-semantic processing (e.g., refs. 70 and 71). In particular, words that occur in the context of phrases or sentences plausibly lead to more robust and longer-lasting representations because they receive reinforcement from related words, which tend to occur in close proximity in natural language (e.g., refs. 72 and 73). Thus, in sentences (compared with lists of unrelated words) the net amount of activation (*i*) should be higher overall and (*ii*) should increase as more words are incorporated into the evolving structure/meaning. Alternatively, these results may reflect the construction of a situation model, an abstract mental representation of a verbally described situation or event (e.g., refs. 74–77). Situation models are typically argued to be multidimensional, with the most common dimensions including protagonist, intentionality, causation, time, and space (e.g., ref. 76). Constructing a coherent situation model has been argued by many to be a signature of successful comprehension of a text (e.g., refs. 74, 75, 78–82). Although most empirical work on situation models has focused on discourse-level understanding, the concept also applies to smaller segments of connected text, like clauses and sentences.

To conclude, the increase in γ-response over the course of the sentence constitutes a neural correlate of the evolving representation of sentence-level meaning. This effect appears to be broadly distributed across the fronto-temporal language network rather than being localized to a particular brain region. The precise nature of this build-up effect remains to be determined, and we lay out several questions whose evaluation would help constrain the space of possible hypotheses.

First, is the build-up effect specific to language, or would other meaningful stimuli (e.g., movies) or structured meaningless stimuli (e.g., music) elicit it as well? The regions of the language network have been shown to respond to language stimuli in a highly selective manner (e.g., refs. 42–44), so it is unlikely that nonlinguistic stimuli would produce a similar build-up effect in these regions (although they might produce such an effect in other cortical regions if similar neural phenomena underlie the representation of complex structured representations in these other domains). However, this possibility remains to be evaluated. Second, is a syntactic frame required or would lists of semantically related words (e.g., cat-dog-pig) or words that could be combined to form a complex meaning (e.g., cookie-girl-eat) suffice? Third, what is the temporal scope of the effect? Clearly, the response cannot increase indefinitely; instead, it will plausibly eventually reach a plateau. The longer time-scale dynamics of the γ build-up remain to be discovered. And fourth, what happens when an incoming word does not fit the structure or meaning constructed so far? From the ERP literature, we know that unexpected words lead to a larger N400 effect (29, 67). However, given that the build-up effect appears to reflect distinct mental computations, it would be important to characterize its behavior with respect to structurally and semantically unexpected words.

In summary, we report here a striking and robust phenomenon: a monotonic increase in neural activity over the course of a sentence as the subject processes it. This effect occurs for sentences, but not for word meanings or syntactic structure alone, ruling out most domain-general accounts and implicating in the effect a quintessential property of language: compositionality. Although these results certainly do not give us a full account of how sentence meaning is constructed and represented neurally, they provide an exciting new window into how that process unfolds over time, and a powerful method for understanding the sequence of neural events that underlie the extraction of complex linguistic meanings in future work.

Materials and Methods

Participants. We recorded electrical activity from intracranial electrodes of six subjects (five female, aged 14–29 y) with intractable epilepsy who read sentences, lists of words, Jabberwocky sentences, and lists of nonwords. These subjects underwent temporary implantation of subdural electrode arrays at Albany Medical College to localize the epileptogenic zones and to delineate it from eloquent cortical areas before brain resection. All of the subjects gave informed written consent to participate in the study, which was approved by the Institutional Review Board of Albany Medical College. Two subjects were excluded from the main analyses because only two electrodes in each subject met our criteria for inclusion in this study. However, their data were qualitatively and quantitatively similar and are included in *SI Appendix, Part D*. One further subject was tested but excluded from all analyses because of difficulties in performing the task (i.e., pressing multiple keys, looking away from the screen) during the first five runs. After the first five runs, the subject required a long break during which a seizure occurred.

Materials and Procedure. In an event-related design, subjects read sentences, lists of words, Jabberwocky sentences, and lists of nonwords. The materials were adapted from ref. 28. Each event (trial) consisted of eight words/nonwords, presented one at a time at the center of the screen. At the end of each sequence, a memory probe was presented (a word in the sentence and word-list conditions, and a nonword in the Jabberwocky and nonword-list conditions) and participants had to decide whether or not the probe had appeared in the preceding sequence by pressing one of two buttons. Two different presentation rates were used: S1, S5, and S6 viewed each word/nonword for 450 ms (fast-timing), and S2, S3, and S4 viewed each word/nonword for 700 ms (slow-timing). The presentation speed was determined before the experiment and was based on the participant's preferences. After the last word/nonword in the sequence, a fixation cross was presented for 250 ms, followed by the probe item (1,400-ms fast-timing, 1,900 ms slow-timing), and a postprobe fixation (250 ms). Behavioral responses were continually recorded. After each trial, a fixation cross was presented for a variable amount of time, semirandomly selected from a range of durations from 0 to 11,000 ms, to obtain a low-level baseline for neural activity.

Trials were grouped into runs to give participants short breaks throughout the experiment. In the fast-timing version of the experiment, each run included eight trials per condition and lasted 220 s, and in the slow-timing version, each run included six trials per condition and lasted 264 s. The total amount of intertrial fixation in each run was 44 s for the fast-timing version and 72 s for the slow-timing version. All subjects completed 10 runs of the experiment, for a total of 80 trials per condition in the fast-timing version and 60 trials per condition in the slow-timing version.

Data Collection and Analysis. The implanted electrode grids consisted of platinum-iridium electrodes that were 4 mm in diameter (2.3–3 mm exposed) and spaced with an interelectrode distance of 0.6 or 1 cm. The total numbers of implanted electrodes were 120, 128, 112, 134, 98, and 36 for the six subjects, respectively (Table 1). Electrodes were implanted on the left hemisphere for all subjects except S6, who had bilateral coverage (16 left hemisphere electrodes). Signals were digitized at 1,200 Hz. Recordings were synchronized with stimulus presentation and stored using the BCI2000 software platform (83, 84). Upon visual inspection of the recordings, we removed reference electrodes, ground, and electrodes with high noise levels and interictal activity as revealed by independent analyses, which left 117, 84, 84, 124, 87, and 33 electrodes for the six subjects, respectively.

Cortical Mapping. We defined the brain anatomy of each subject using preoperative MRI scans, and the location of the electrodes using postoperative computed tomography (CT) imaging. We then created a 3D surface model of each subject's cortex from the MRI images, coregistered it with the location of the electrodes given by the CT images using Curry Software (Compumedics NeuroScan).

Extraction of ECoG Signal Envelope. ECoG recordings were first high-pass–filtered at a frequency of 0.5 Hz and spatially distributed noise common to all electrodes was removed using a common average reference spatial filter. Notch filters removed further noise at 60, 120, 180, and 240 Hz. An IIR bandpass filter was used to select high γ-frequencies (70–170 Hz) and the envelope of the ECoG signal (i.e., the magnitude of the analytic signal) in the high γ-band was computed by taking the absolute value of the Hilbert transform of the resulting signal. The signal envelopes were further low-pass–filtered at 100 Hz and down-sampled to 300 Hz to reduce noise.

Selecting EOIs.

Step 1: Selection of language-responsive electrodes. Language-responsive electrodes were defined as electrodes in which the envelope of the high γ-signal is significantly higher for trials of the sentence condition than the nonword-list condition. To do this, we first computed the mean of the signal envelope for each of the eight word positions (time-locked to the onset of each word/nonword and averaging over the presentation window) in each trial for each condition in each electrode, using data from the odd-numbered runs only (see *SI Appendix, Part F* for sample EOIs without within-position averaging). We then computed the mean across the eight word positions in each trial for each condition in each electrode. Finally, we correlated the trial means with a vector of condition labels (sentences = 1, nonword-lists = −1). The resulting Spearman's ρ provided a benchmark against which to test the significance of any positive correlations. The condition labels vector was randomly reordered (via a permutation test without replacement) and a new Spearman's ρ was computed, and this process was repeated 1,000 times. The fraction of correlations from randomly assigned labels that produced a higher ρ than the benchmark correlation became our P value. Electrodes with $P \leq 0.01$ and a positive ρ were included in step 2.

Step 2: Selection of the subset of language-responsive electrodes exhibiting an increase across word positions in the sentence condition. These electrodes were defined as electrodes that exhibited a monotonic increase in high γ-signal over the course of a sentence. To do this, we computed the mean of the high γ-envelope for each word position in the sentence condition and then computed the Spearman's rank correlation coefficient, which measures statistical dependence without assuming linearity, between word position (one through eight) and mean signal magnitude at each position within each sentence trial, separately for each electrode. As in step 1, we again only used the data from odd-numbered runs. A two-tailed one-sample t test was performed and electrodes with mean correlations that were significantly different from zero ($P \leq 0.01$) were selected as our EOIs and were included in all of the analyses reported here.

ACKNOWLEDGMENTS. We thank Ted Gibson, Charles Jennings, Roger Levy, Kyle Mahowald, Steve Piantadosi, and Nathaniel Smith for providing helpful comments on this work; Eyal Dechter for help with setting up the experiment; Steve Piantadosi and Kyle Mahowald for providing some word-level features extracted from corpora; and Zuzanna Balewski and Ted Gibson for help with collecting norming data on Mechanical Turk. This research was supported by the NIH (Grants EB00856, EB006356, and EB018783), the US Army Research Office (Grants W911NF-08-1-0216, W911NF-12-1-0109, W911NF-14-1-0440), and Fondazione Neurone. E.F. was supported by National Institute of Child Health and Human Development Award HD-057522.

1. Geschwind N (1970) The organization of language and the brain. *Science* 170(3961):940–944.
2. Binder JR, et al. (1997) Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci* 17(1):353–362.
3. Bates E, et al. (2003) Voxel-based lesion-symptom mapping. *Nat Neurosci* 6(5):448–450.
4. Jacobs J, Kahana MJ (2010) Direct brain recordings fuel advances in cognitive electrophysiology. *Trends Cogn Sci* 14(4):162–171.
5. Mukamel R, Fried I (2012) Human intracranial recordings and cognitive neuroscience. *Annu Rev Psychol* 63:511–537.
6. Chang EF, et al. (2010) Categorical speech representation in human superior temporal gyrus. *Nat Neurosci* 13(11):1428–1432.
7. Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495(7441):327–332.
8. Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF (2013) Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc Natl Acad Sci USA* 110(7):2653–2658.
9. Bouchard KE, Chang EF (2014) Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. *J Neurosci* 34(38):12662–12677.

10. Cogan GB, et al. (2014) Sensory-motor transformations for speech occur bilaterally. *Nature* 507(7490):94–98.
11. Leonard MK, Bouchard KE, Tang C, Chang EF (2015) Dynamic encoding of speech sequence probability in human temporal cortex. *J Neurosci* 35(18):7203–7214.
12. Hart J, Jr, et al. (1998) Temporal dynamics of verbal object comprehension. *Proc Natl Acad Sci USA* 95(11):6498–6503.
13. Tanji K, Suzuki K, Delorme A, Shamoto H, Nakasato N (2005) High-frequency gamma-band activity in the basal temporal cortex during picture-naming and lexical-decision tasks. *J Neurosci* 25(13):3287–3293.
14. Mainy N, et al. (2008) Cortical dynamics of word recognition. *Hum Brain Mapp* 29(11): 1215–1230.
15. Towle VL, et al. (2008) ECoG gamma activity during a language task: Differentiating expressive and receptive speech areas. *Brain* 131(Pt 8):2013–2027.
16. Sahin NT, Pinker S, Cash SS, Schomer D, Halgren E (2009) Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* 326(5951):445–449.
17. Edwards E, et al. (2010) Spatiotemporal imaging of cortical activation during verb generation and picture naming. *NeuroImage* 50(1):291–301.
18. Kellis S, et al. (2010) Decoding spoken words using local field potentials recorded from the cortical surface. *J Neural Eng* 7(5):056007.
19. Thampratankul L, et al. (2010) Cortical gamma oscillations modulated by word association tasks: Intracranial recording. *Epilepsy Behav* 18(1-2):116–118.
20. Flinker A, Chang EF, Barbaro NM, Berger MS, Knight RT (2011) Sub-centimeter language organization in the human temporal lobe. *Brain Lang* 117(3):103–109.
21. Pei X, Barbour DL, Leuthardt EC, Schalk G (2011) Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *J Neural Eng* 8(4):046028.
22. Conner CR, Chen G, Pieters TA, Tandon N (2014) Category specific spatial dissociations of parallel processes underlying visual naming. *Cereb Cortex* 24(10):2741–2750.
23. Kojima K, et al. (2013) Gamma activity modulated by picture and auditory naming tasks: Intracranial recording in patients with focal epilepsy. *Clin Neurophysiol* 124(9): 1737–1744.
24. Trébuchon A, Démonet JF, Chauvel P, Liégeois-Chauvel C (2013) Ventral and dorsal pathways of speech perception: An intracerebral ERP study. *Brain Lang* 127(2):273–283.
25. Cibelli ES, Leonard MK, Johnson K, Chang EF (2015) The influence of lexical statistics on temporal lobe cortical dynamics during spoken word listening. *Brain Lang* 147:66–75.
26. Flinker A, et al. (2015) Redefining the role of Broca's area in speech. *Proc Natl Acad Sci USA* 112(9):2871–2875.
27. von Humboldt W (1836) On language. *On the Diversity of Human Language Construction and Its Influence on the Mental Development of the Human Species*, ed Losonsky M (Cambridge Univ Press, New York).
28. Fedorenko E, Hsieh PJ, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N (2010) New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *J Neurophysiol* 104(2):1177–1194.
29. Kutas M, Hillyard SA (1980) Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207(4427):203–205.
30. Osterhout L, Holcomb PJ (1992) Event-related brain potentials elicited by syntactic anomaly. *J Mem Lang* 31(6):785–806.
31. Hagoort P, Brown C, Groothusen J (1993) The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Lang Cogn Process* 8(4):439–483.
32. Hagoort P (2008) The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philos Trans R Soc Lond B Biol Sci* 363(1493): 1055–1069.
33. Van Petten C, Luka BJ (2012) Prediction during language comprehension: Benefits, costs, and ERP components. *Int J Psychophysiol* 83(2):176–190.
34. Kaan E (2007) Event-related potentials and language processing: A brief overview. *Lang Linguist Compass* 1(6):571–591.
35. Kielar A, Panamsky L, Links KA, Meltzer JA (2015) Localization of electrophysiological responses to semantic and syntactic anomalies in language comprehension with MEG. *NeuroImage* 105:507–524.
36. Halgren E, et al. (2002) N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage* 17(3): 1101–1116.
37. Kuperberg G, et al. (2003) Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *J Cogn Neurosci* 15(2):272–293.
38. Bavelier D, et al. (1998) Hemispheric specialization for English and ASL: Left invariance-right variability. *NeuroReport* 9(7):1537–1542.
39. Robertson DA, et al. (2000) Functional neuroanatomy of the cognitive process of mapping during discourse comprehension. *Psychol Sci* 11(3):255–260.
40. Noppeney U, Price CJ (2004) An FMRI study of syntactic adaptation. *J Cogn Neurosci* 16(4):702–713.
41. Snijders TM, et al. (2009) Retrieval and unification of syntactic structure in sentence comprehension: An FMRI study using word-category ambiguity. *Cereb Cortex* 19(7): 1493–1503.
42. Fedorenko E, Behr MK, Kanwisher N (2011) Functional specificity for high-level linguistic processing in the human brain. *Proc Natl Acad Sci USA* 108(39):16428–16433.
43. Monti MM, Parsons LM, Osherson DN (2009) The boundaries of language and thought in deductive inference. *Proc Natl Acad Sci USA* 106(30):12554–12559.
44. Monti MM, Parsons LM, Osherson DN (2012) Thought beyond language: Neural dissociation of algebra and natural language. *Psychol Sci* 23(8):914–922.
45. Braze D, et al. (2011) Unification of sentence processing via ear and eye: An fMRI study. *Cortex* 47(4):416–431.
46. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: The dangers of double dipping. *Nat Neurosci* 12(5):535–540.
47. Manning JR, Jacobs J, Fried I, Kahana MJ (2009) Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *J Neurosci* 29(43):13613–13620.
48. Miller KJ, Sorensen LB, Ojemann JG, den Nijs M (2009) Power-law scaling in the brain surface electric potential. *PLoS Comput Biol* 5(12):e1000609.
49. Ray S, Maunsell JH (2011) Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol* 9(4):e1000610.
50. Whittingstall K, Logothetis NK (2009) Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. *Neuron* 64(2):281–289.
51. Bedny M, Pascual-Leone A, Dodell-Feder D, Fedorenko E, Saxe R (2011) Language processing in the occipital cortex of congenitally blind adults. *Proc Natl Acad Sci USA* 108(11):4429–4434.
52. Pallier C, Devauchelle AD, Dehaene S (2011) Cortical representation of the constituent structure of sentences. *Proc Natl Acad Sci USA* 108(6):2522–2527.
53. Fedorenko E, Nieto-Castañon A, Kanwisher N (2012) Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia* 50(4):499–513.
54. Friederici AD, Bahlmann J, Heim S, Schubotz RI, Anwander A (2006) The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proc Natl Acad Sci USA* 103(7):2458–2463.
55. Grodzinsky Y, Friederici AD (2006) Neuroimaging of syntax and syntactic processing. *Curr Opin Neurobiol* 16(2):240–246.
56. Hagoort P (2005) On Broca, brain, and binding: A new framework. *Trends Cogn Sci* 9(9):416–423.
57. Vandenberghe R, Nobre AC, Price CJ (2002) The response of left temporal cortex to sentences. *J Cogn Neurosci* 14(4):550–560.
58. Brennan J, et al. (2012) Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain Lang* 120(2):163–173.
59. Bemis DK, Pylkkänen L (2013) Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cereb Cortex* 23(8):1859–1873.
60. Baron SG, Osherson D (2011) Evidence for conceptual combination in the left anterior temporal lobe. *NeuroImage* 55(4):1847–1852.
61. Blank I, Balewski Z, Mahowald K, Fedorenko E (2016) Syntactic processing is distributed across the language system. *NeuroImage* 127:307–323.
62. Bautista A, Wilson SM (2016) Neural responses to grammatically and lexically degraded speech. *Lang Cogn Neurosci* 31(4):567–574.
63. Blank I, Kanwisher N, Fedorenko E (2014) A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *J Neurophysiol* 112(5):1105–1118.
64. Fedorenko E, Duncan J, Kanwisher N (2013) Broad domain generality in focal regions of frontal and parietal cortex. *Proc Natl Acad Sci USA* 110(41):16616–16621.
65. Marslen-Wilson W, Tyler LK (1975) Processing structure of sentence perception. *Nature* 257(5529):784–786.
66. Levy R (2008) Expectation-based syntactic comprehension. *Cognition* 106(3):1126–1177.
67. Kutas M, Federmeier KD (2011) Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol* 62:621–647.
68. Graesser AC (2013) *Prose Comprehension Beyond the Word* (Springer Science & Business Media, Medford, MA).
69. Smith NJ, Levy R (2013) The effect of word predictability on reading time is logarithmic. *Cognition* 128(3):302–319.
70. Quillian MR (1969) The teachable language comprehender: A simulation program and theory of language. *Commun ACM* 12(8):459–476.
71. Collins AM, Loftus EF (1975) A spreading-activation theory of semantic processing. *Psychol Rev* 82(6):407–428.
72. Lund K, Burgess C (1996) Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav Res Methods Instrum Comput* 28(2):203–208.
73. Landauer TK, Dumais ST (1997) A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol Rev* 104(2):211–240.
74. Johnson-Laird PN (1983) *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness* (Harvard Univ Press, Cambridge, MA).
75. van Dijk TA, Kintsch W (1983) *Strategies in Discourse Comprehension* (Academic, New York).
76. Gernsbacher MA (1990) *Language Comprehension as Structure Building* (Erlbaum, Hillsdale, NJ).
77. Zwaan RA, Radvansky GA (1998) Situation models in language comprehension and memory. *Psychol Bull* 123(2):162–185.
78. Johnson-Laird PN (1989) Mental models. *Foundations of Cognitive Science*, ed Posner MI (MIT Press, Cambridge, MA), pp 469–499.
79. Perfetti CA (1989) There are generalized abilities and one of them is reading. *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, ed Resnick LB (Erlbaum, Hillsdale, NJ), pp 307–335.
80. Graesser AC, Singer M, Trabasso T (1994) Constructing inferences during narrative text comprehension. *Psychol Rev* 101(3):371–395.
81. Zwaan RA, Langston MC, Graesser AC (1995) The construction of situation models in narrative comprehension: An event-indexing model. *Psychol Sci* 6:292–297.
82. Graesser AC, Millis KK, Zwaan RA (1997) Discourse comprehension. *Annu Rev Psychol* 48:163–189.
83. Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR (2004) BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Trans Biomed Eng* 51(6): 1034–1043.
84. Schalk G, Mellinger JA (2010) *Practical Guide to Brain-Computer Interfacing with BCI2000* (Springer, London).

Supplemental Information for Fedorenko et al. "A neural correlate of the construction of sentence meaning"

SI Table of Contents:

a. Evaluating the significance of the build-up effect (i.e., of the correlations between signal magnitude and word position within the trial) in each condition in the four subjects, estimated in data from even-numbered runs (i.e., data not used for selecting the EOIs).

	Electrode Number	Sentences	Word-lists	Jabberwocky	Nonword-lists
Subject 1	21	0.54	0.06	0.14	-0.06
	22	0.35	0.04	0.16	0.06
	26	0.51	0.18	0.18	-0.08
	28	0.27	0.05	0.00	-0.14
	33	0.31	0.02	0.14	-0.04
	34	0.37	0.08	0.15	0.01
	45	0.35	0.04	0.06	-0.11
	54	0.34	0.09	-0.13	-0.24
	58	0.21	0.09	-0.01	-0.19
	59	0.27	0.00	-0.02	-0.12
	60	0.37	0.15	0.12	-0.13
	63	0.33	0.11	-0.08	-0.05
	64	0.46	0.06	0.00	-0.14
	67	0.43	0.19	0.21	0.03
	79	0.48	0.06	0.12	0.05
	80	0.52	0.07	-0.03	-0.08
	83	0.34	0.04	-0.01	-0.13
	84	0.32	0.03	-0.02	0.07
	87	0.31	0.00	0.09	-0.05
	88	0.29	-0.03	-0.04	-0.04
	90	0.35	0.00	0.07	0.02
	91	0.33	0.13	-0.02	-0.07
	92	0.42	0.05	0.14	0.03
	96	0.32	-0.08	0.03	-0.10
	100	0.43	0.07	0.14	-0.19
	101	0.21	-0.10	-0.01	-0.11
	102	0.25	-0.02	0.01	-0.12
	Across EOIs	**0.36**	**0.05**	**0.05**	**-0.07**
Subject 3	12	0.29	0.25	0.15	0.18
	34	0.34	0.10	0.05	0.06
	36	0.12	0.03	-0.04	-0.10
	45	0.12	0.10	-0.05	-0.13
	49	0.17	0.11	0.01	-0.06
	59	0.34	0.05	0.03	-0.03
	60	0.39	0.12	0.05	-0.06
	86	0.31	0.09	0.08	-0.03
	94	0.23	-0.02	0.03	-0.11
	Across EOIs	**0.26**	**0.09**	**0.03**	**-0.03**
Subject 4	3	0.21	-0.05	-0.10	0.04
	4	0.30	0.14	-0.04	-0.03
	10	0.19	-0.03	-0.12	-0.07
	18	0.42	0.16	-0.03	0.02
	19	0.33	0.14	-0.03	0.10
	20	0.16	0.05	0.00	0.09
	38	0.33	0.06	0.01	-0.10
	39	0.28	-0.01	0.09	-0.20
	101	0.44	0.11	-0.06	-0.04

	Across EOIs	**0.29**	**0.06**	**-0.03**	**-0.02**
Subject 5	12	0.29	0.26	0.01	-0.08
	24	0.29	0.10	0.21	0.01
	75	0.07	-0.10	-0.07	0.09
	82	0.15	-0.03	0.00	0.00
	87	0.10	0.09	0.11	-0.11
	91	0.42	0.14	0.02	-0.03
	Across EOIs	**0.22**	**0.08**	**0.05**	**-0.02**
Across subjects		**0.28**	**0.07**	**0.03**	**-0.04**

Table SI-1. Spearman's rho computed between signal magnitude and word position within the trial for each subject and each EOI and averaged across EOIs and subjects, in each experimental condition. Two-tailed one-sample *t*-tests were performed to test if the correlations were significantly different from zero. Significance levels are shaded with white: $p < 0.01$, mid-gray: $0.01 <= p < 0.05$, and dark-gray: $p >= 0.05$. Positive correlations indicate that signal magnitude increases for later word positions.

	Position 1	Position 2	Position 3	Position 4	Position 5	Position 6	Position 7	Position 8
Subject 1	-2.18 ± 0.95 $t(26) = -2.29$ $p = 0.03$	0.76 ± 1.01 $t(26) = 0.75$ $p = 0.46$	5.18 ± 1.94 $t(26) = 2.67$ $p = 0.01$	12.53 ± 2.50 $t(26) = 5.01$ $p < 10^{-4}$	14.71 ± 2.40 $t(26) = 6.12$ $p < 10^{-5}$	17.35 ± 2.49 $t(26) = 6.97$ $p < 10^{-6}$	21.73 ± 2.51 $t(26) = 8.66$ $p < 10^{-8}$	26.79 ± 2.45 $t(26) = 10.92$ $p < 10^{-10}$
Subject 3	-3.88 ± 1.42 $t(8) = -2.74$ $p = 0.03$	-0.88 ± 1.46 $t(8) = -0.60$ $p = 0.56$	5.69 ± 0.90 $t(8) = 6.35$ $p < 10^{-3}$	10.28 ± 1.43 $t(8) = 7.18$ $p < 10^{-4}$	11.32 ± 1.43 $t(8) = 7.90$ $p < 10^{-4}$	9.59 ± 1.99 $t(8) = 4.82$ $p < 10^{-2}$	6.10 ± 2.49 $t(8) = 2.45$ $p = 0.04$	15.86 ± 2.12 $t(8) = 7.48$ $p < 10^{-4}$
Subject 4	1.03 ± 0.59 $t(8) = 1.73$ $p = 0.12$	3.35 ± 0.99 $t(8) = 3.38$ $p = 0.01$	6.68 ± 1.19 $t(8) = 5.63$ $p < 10^{-3}$	9.96 ± 2.21 $t(8) = 4.51$ $p < 10^{-2}$	11.49 ± 2.07 $t(8) = 5.56$ $p < 10^{-3}$	13.43 ± 1.56 $t(8) = 8.60$ $p < 10^{-4}$	11.66 ± 1.91 $t(8) = 6.12$ $p < 10^{-3}$	16.11 ± 2.38 $t(8) = 6.77$ $p < 10^{-3}$
Subject 5	-0.31 ± 1.00 $t(5) = -0.32$ $p = 0.77$	2.66 ± 2.33 $t(5) = 1.14$ $p = 0.30$	5.05 ± 2.35 $t(5) = 2.15$ $p = 0.08$	9.90 ± 4.43 $t(5) = 2.24$ $p = 0.08$	8.78 ± 3.91 $t(5) = 2.24$ $p = 0.07$	9.84 ± 4.81 $t(5) = 2.05$ $p = 0.10$	11.63 ± 4.52 $t(5) = 2.57$ $p = 0.05$	13.11 ± 6.45 $t(5) = 2.03$ $p = 0.10$

Table SI-2. The average differences between *sentences* and *nonwords* were computed across EOIs for each subject at each position in the sequence. Standard errors are reported. Two-tailed paired sample *t*-tests were used. Significance levels are shaded with white: $p < 0.01$, mid-gray: $0.01 <= p < 0.05$, and dark-gray: $p >= 0.05$.

---

b. Evidence that similar results obtain when using even-numbered runs for selecting the EOIs and odd-numbered runs for examining the significance of the results.

As shown in Table SI-3, there is substantial overlap between the EOIs selected based on odd-numbered vs. even-numbered runs. Moreover, Figure SI-1 demonstrates that the pattern of results is virtually identical to the one reported in the main analysis (Figure 2a-b).

	Subject 1	Subject 3	Subject 4	Subject 5
EOIs odd	27	9	9	6

EOIs even	31	6	8	5
Intersection	21	4	6	4
% of odd-runs EOIs that are also even-runs EOIs	77.78%	44.44%	66.67%	66.67%

Table SI-3. Overlap between EOIs defined based on data from odd-numbered runs (as in our main analysis) vs. even-numbered runs.

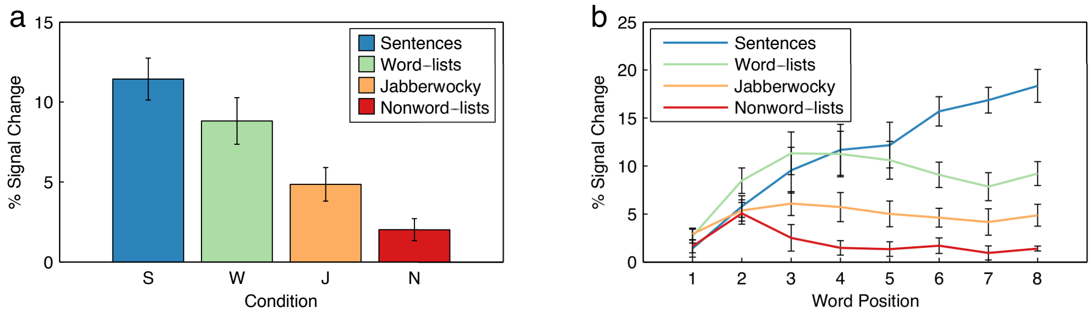

Figure SI-1. a. Gamma magnitude for all conditions (S=sentences, W=word-lists, J=Jabberwocky sentences, N=nonword-lists) averaged across word positions and subjects estimated in odd-numbered runs (with the EOIs being selected using data from even-numbered runs).
b. Gamma magnitude for all conditions averaged across subjects in each of eight word positions in odd-numbered runs.
Error bars indicate standard errors of the mean over subjects in both panels.

c. Behavioral (memory-probe task) performance.

	Sentences	Word-lists	Jabberwocky	Nonword-lists
Subject 1	96.25%	91.25%	80.00%	78.48%
Subject 3	100%	91.67%	86.67%	71.67%
Subject 4	100%	93.33%	90.00%	78.33%
Subject 5*	45.00%	65.00%	48.75%	51.25%

Table SI-4. Percent accuracy on the behavioral task across all runs of the experiment.

* Subject 5 performed poorly across conditions due to misunderstanding the instructions. This subject answered “yes” for all sentence trials and almost always answered “no” for all Jabberwocky and nonword-list trials. However, because a) ECoG data are highly valuable, b) language comprehension was shown to produce similar activations in fMRI regardless of whether a memory-probe task was included [1], and c) this subject’s ECoG data patterns looked qualitatively similar to those of the other subjects, we included this subject’s data in the analyses.

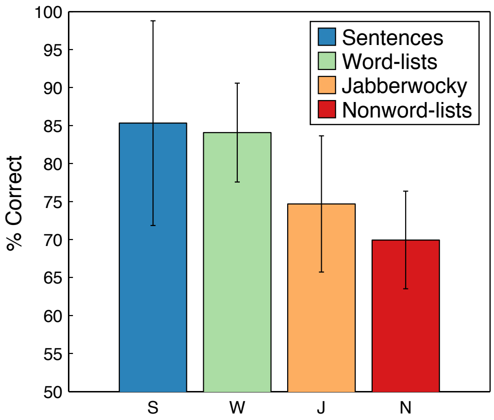

Figure SI-2. Task accuracy averaged across the four subjects. Error bars indicate standard errors of the mean over subjects.

No significant differences were found between subjects' accuracies during the sentence and word-list conditions (t(3) = 0.17, p = 0.88), between the sentence and Jabberwocky conditions (t(3) = 2.04, p = 0.13), or between the sentence and nonword-list conditions (t(3) = 2.04, p = 0.13).

A typical accuracy-RT relationship was observed, such that the conditions with lower accuracy rates were also the ones with longer RTs. In particular, averaging across the four subjects, the means and standard errors were as follows: *sentences*: 879 ± 118ms, *word-lists*: 932 ± 68ms, *jabberwocky*: 920 ± 129ms, and *nonword-lists*: 973 ± 131ms.

d. Results from the two subjects excluded from the main analyses due to a small number of EOIs.

	Total Electrodes	Analyzed Electrodes	S>N Electrodes	S>N and S Increasing (i.e.,EOIs)
Subject 2	128	84	6 (0.07*)	2 (0.02**, 0.33***)
Subject 6	36	33	3 (0.09)	2 (0.06, 0.67)

Table SI-5. Details on the numbers of electrodes in the two subjects excluded from the main analyses.
*proportion of *S>N* electrodes relative to analyzed electrodes
**proportion of *S>N AND S Increasing* electrodes relative to analyzed electrodes
***proportion of *S>N AND S Increasing* electrodes relative to *S>N* electrodes

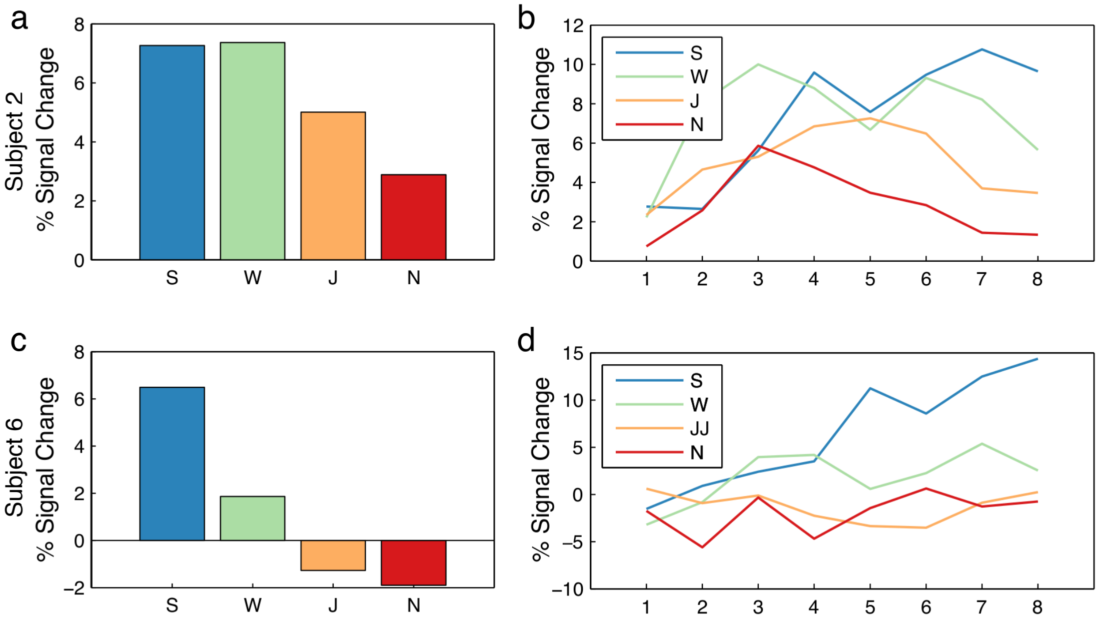

Figure SI-3. a. High gamma magnitude for all conditions (S=sentences, W=word-lists, J=Jabberwocky sentences, N=nonword-lists) averaged across word positions estimated in even-numbered runs, from excluded subjects.
b. High gamma magnitude for all conditions averaged across EOIs in each of eight word positions in even-numbered runs, from excluded subjects.

	S - N	S - W	S - J	W - N	J - N	W - J
Subject 2	4.38 ± 2.69 t(1) = 6.80 p = 0.35	-0.10 ± 2.43 t(1) = 10.67 p = 0.97	2.26 ± 2.63 t(1) = 11.21 p = 0.55	4.48 ± 0.26 t(1) = 3.08 p = 0.04	2.13 ± 0.06 t(1) = 0.57 p = 0.02	2.36 ± 0.20 t(1) = 5.65 p = 0.05
Subject 6	8.39 ± 3.17 t(1) = 6.58 p = 0.22	4.63 ± 3.00 t(1) = 4.18 p = 0.37	7.77 ± 3.81 t(1) = 6.71 p = 0.29	3.76 ± 0.16 t(1) = 4.67 p = 0.03	0.63 ± 0.65 t(1) = 2.43 p = 0.51	3.14 ± 0.81 t(1) = 7.38 p = 0.16

Table SI-6. Excluded subjects' results from two-tailed paired samples *t*-tests on the differences in percent signal change (PSC, with respect to baseline fixation) between pairs of experimental conditions. Data is taken from even-numbered runs only and averaged across all word positions and electrodes of interest (EOIs). The first row of each cell denotes the average PSC difference with standard errors of the mean across EOIs. Significance levels are shaded with white: $p < 0.01$, mid-gray: $0.01 \leq p < 0.05$, and dark-gray: $p \geq 0.05$.

	Channel Number	S	W	J	N
Subject 2	36	0.20	0.05	0.16	0.06
	52	0.25	0.05	-0.14	-0.10
	Across EOIs	0.22	0.05	0.01	-0.02
Subject 6	3	0.11	0.08	-0.03	0.02
	22	0.35	0.12	0.01	0.18
	Across EOIs	0.23	0.10	-0.01	0.10

Table SI-7. Excluded subjects' Spearman's rho computed between signal magnitude and word position within the trial for each subject and each EOI and averaged across EOIs and subjects, in each experimental condition. Two-tailed one-sample *t*-tests were performed to test if the correlations were significantly different from zero. Significance levels are shaded with white: p < 0.01, mid-gray: 0.01 <= p < 0.05, and dark-gray: p >= 0.05. Positive correlations indicate that signal magnitude increases for later word positions.

	S	W	J	N
Subject 2	98.33%	91.67%	86.67%	71.67%
Subject 6	90.00%	86.25%	73.75%	65.00%

Table S1-8. Excluded subjects' percent accuracy on the behavioral task across all runs of the experiment.

e. Evaluating the relationship between the build-up effect and changes in i) the strength of predictions for upcoming material, or ii) the predictability of incoming words, across the sentence.

To test whether the sentence build-up effect may reflect the strengthening of one's predictions about the upcoming material over the course of a sentence (e.g., [2]), we took advantage of the fact that sentences differed with respect to their structure and thus with respect to the point where the sentence could end (conveying a complete thought). For example, although a sentence like BY TELLING THE TRUTH WE HURT MATTHEW'S FEELINGS is not complete until the readers see the last word, other sentences (e.g., THE DOG CHASED THE CAT ALL DAY LONG) can end earlier (e.g., at CAT), with the subsequent material being semantically and syntactically optional. For each sentence, we identified the earliest point at which the sentence could be complete. We then averaged the ECoG responses across trials – aligned with respect to their earliest completeness point – and across EOIs within each subject, and then across subjects. Sentences that were not judged to be complete until the eighth word were not included in this analysis. This left 29, 22, 22, and 29 sentence trials taken from even-numbered runs for subjects 1, 3, 4, and 5 respectively. We then examined the high gamma magnitude at the word that could complete the sentence (word n) and the following word (word n+1) relative to the two preceding words (words n-2 and n-1). If the build-up effect reflects the *strength of one's predictions* about the upcoming material, there should be a drop in signal at word n, because at that point expectations for upcoming material weaken significantly (given that the sentence could end at this point).

Alternatively, the build-up effect could reflect the *predictability* of incoming words (which can be formalized in terms of *surprisal*, i.e., the negative log-probability of a word given the context [2]), with stronger responses for words that satisfy predictions formed based on the preceding context. In that case, the signal should drop at word n+1 given that at that point surprisal should be quite high. As noted in the main text, however, this possibility does not seem a priori likely: it is unclear why neural activity would *increase* for words that are easier to integrate into the preceding context, especially given that in both behavioral measures (e.g., [3]) and other neural measures (e.g., [4]), higher responses have been reported for *less predictable / more surprising* words.

The results, shown in Figure SI-4, did not reveal a drop on word *n*; instead, the signal continues to increase, suggesting that the build-up effect probably does not reflect the expectations that increase in strength as the sentence unfolds. Although there was a small numerical decrease on word *n*+1, the signal was not significantly lower than on word *n* (either across subjects: 13.86 vs. 12.82, t(3) = 0.87, p = 0.45; or across trials, averaging across all EOIs from the four subjects: 15.70 vs. 15.19, t(50) = 0.51, p = 0.61), and the size of the drop is not commensurate with the change in surprisal levels.

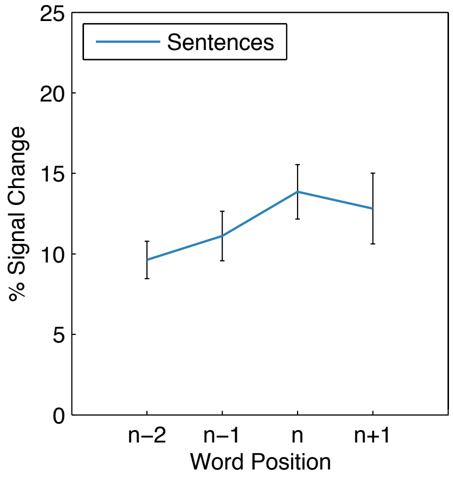

Figure SI-4. Percent signal change (PSC), relative to the baseline fixation, in the sentence condition for word *n* (a word at which the sentence could be complete), word *n*+1 (a word after the point at which the sentence could be complete), and two preceding words. Error bars indicate standard errors of the mean over subjects.

In addition to examining the pattern of gamma activity at and after a possible point of syntactic and semantic completeness averaging across trials and EOIs, we examined individual trials and individual EOIs to ensure that we are not missing any important data patterns that could shed light on the interpretation of the build-up effect. Although some EOIs do show decreases at or after a possible point of sentence completeness on some trials, none appear to show this pattern reliably across trials. Figure SI-5 shows the responses for four sample trials in each of eight word positions for each of the four subjects. Note that the data are necessarily noisy given that these are single trial data. The noisiness of these single-trial time-courses underlines the importance of investigating their test-retest reliability, as we did in the current manuscript by using a split-half approach.

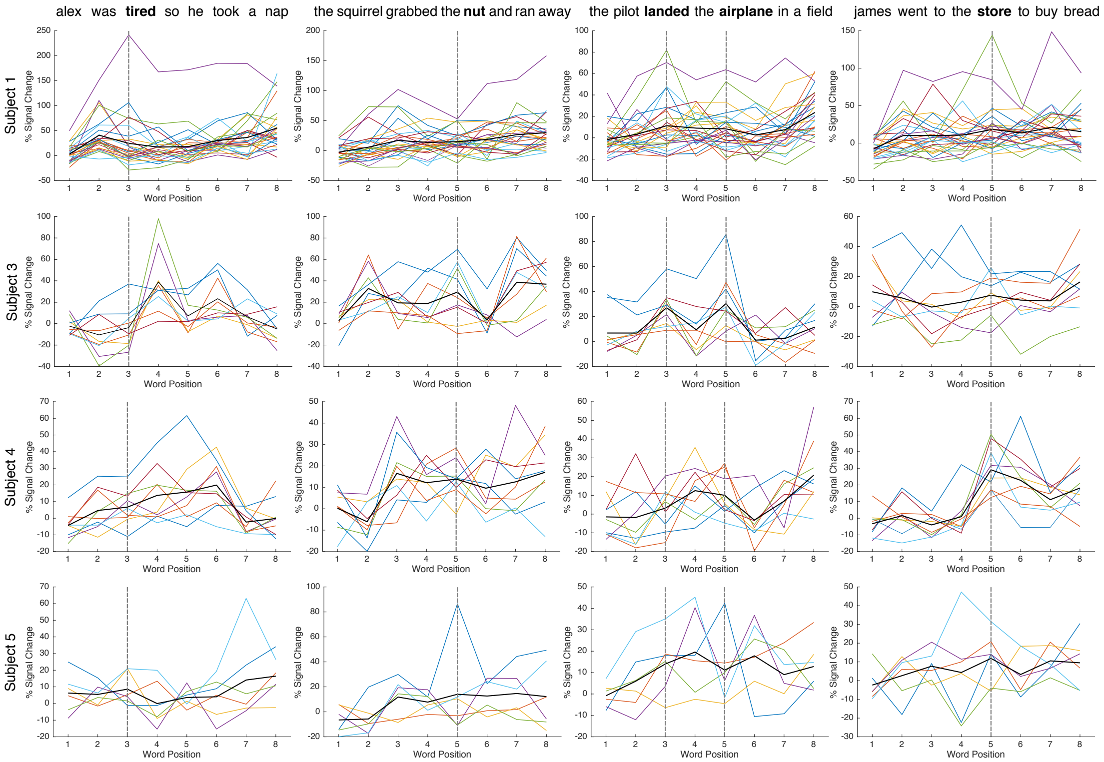

Figure SI-5. Percent signal change (PSC), relative to the baseline fixation, in each of eight word positions for four sample trials (columns) for each of the four subjects (rows). The words that correspond to the points of syntactic and semantic completeness are marked in bold, and the positions of these words are marked with vertical dotted lines. Each graph shows the responses of each individual EOI (in different colors) as well as the average response (shown in a thicker black line).

In summary, the build-up effect is not likely tracking surprisal levels, although a definitive answer will be provided in future work by examining a more structurally diverse set of materials.

f. Data from sample subjects and EOIs shown without within-position averaging

Figure SI-6 shows time-courses for two sample subjects (two EOIs each) for the critical – sentences and nonword-lists – conditions. Here, we did not average gamma within each word position (cf. Figures 1-2 and 4 in the main text) in order to examine the time-courses in greater temporal detail. Some electrodes appear to show a response phase-locked to each word/nonword in the sentence/list, in addition to the orthogonal effect reported here of a gradual build-up across the sentence.

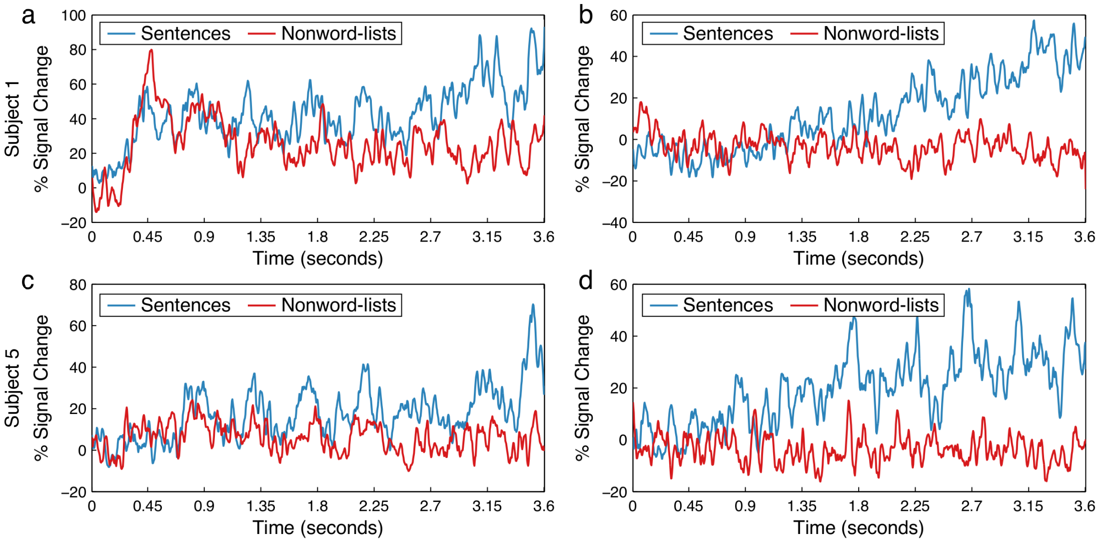

Figure SI-6. Gamma magnitude for the sentence and nonword-list conditions in two sample subjects with two sample EOIs each in even-numbered runs.

---

g. Exploring potential anatomical differences between frontal, lateral temporal, and ventral temporal EOIs

(See the main text for the discussion.)

	Subject 1	Subject 3	Subject 4	Subject 5	Total Contributing Subjects
Left frontal	8	3	7	3	4
Left lateral temporal	12	3	2	3	4
Left ventral temporal	6	3	0	0	2

Table SI-9. Numbers of EOIs in each group by subject.

In addition to examining the differences between frontal, lateral temporal, and ventral temporal EOIs, we examined potential differences across the lateral temporal surface, contrasting EOIs located in the anterior vs. posterior portions. Results, shown in Figure SI-7, reveal broadly similar patterns both in the pattern of mean responses (averaged across positions) and in the time-courses. The only qualitative difference is that the sentence condition appears to separate from the word-list condition earlier in the posterior EOIs, suggesting perhaps that meaning/structure composition occurs earlier in those sites. However, given the small number of EOIs, it is difficult to determine the meaningfulness of this difference. Thus, although there are likely to be functional differences in the language-responsive regions located across the lateral temporal surface, they do not manifest strongly in the mean responses or the time-courses for our four conditions in the current study, perhaps because of the relatively small numbers of EOIs and/or sparse coverage. We plan to pursue this question in future work by using high-density electrode grids.

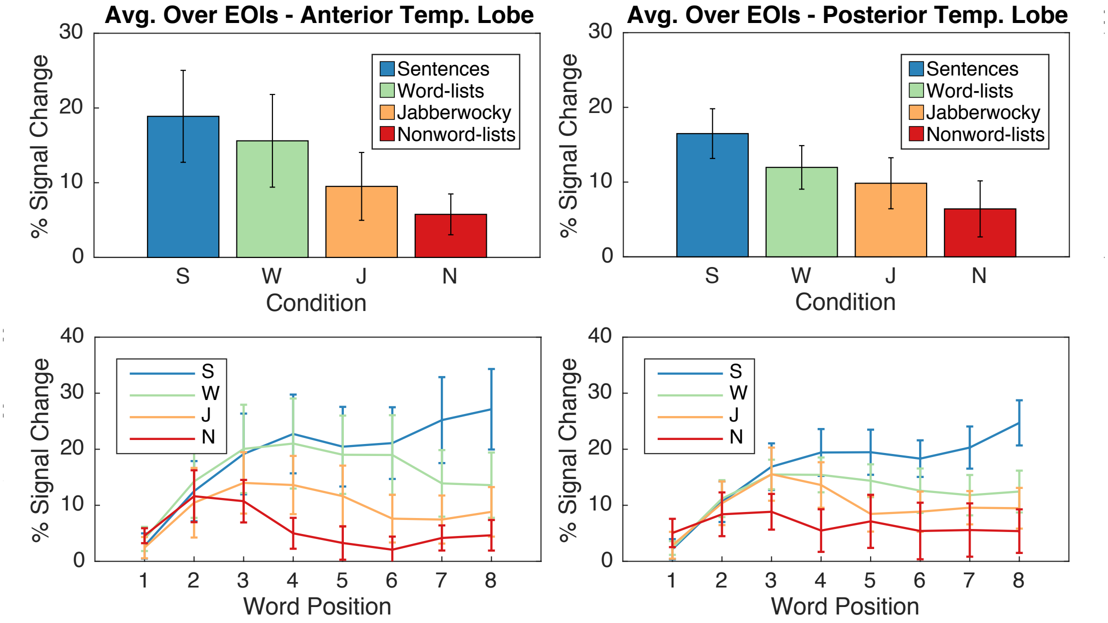

Figure SI-7. EOIs from the four main subjects were further separated into anterior lateral temporal (n = 11) and posterior lateral temporal (n = 5) groups and averaged together to find percent signal change for each condition across positions (top) and in each of the eight positions (bottom). Error bars represent standard errors over EOIs.

---

h. Generalizing the key results to auditory presentation modality in three new subjects

Auditory Control Experiment

High-level language processing regions in the frontal and temporal cortices should reflect a deep analysis of the linguistic content of the stimuli, not a response to their surface form. Indeed, prior neuroimaging studies have found a higher response to sentences versus control stimuli in these regions for both written and spoken materials (e.g., [1,5]). Here we tested whether the same modality independence was true for the build-up effect reported here.

Results

Three patients read or listened to sentences, word-lists, Jabberwocky sentences, and nonword-lists. We used the data from the visual trials to define electrodes of interest (EOIs; see Experimental Procedures below) and then examined the responses during the auditory trials. In these three patients (Subjects A1, A2 and A3), we identified 12, 7, and

1 EOIs, respectively. The results are shown in Figure SI-8; the behavioral data are reported in Table SI-10. As expected given prior fMRI findings of modality-independence of the high-level language processing regions, we observed a pattern that is very similar to the pattern in our original visual-presentation experiment (Figure 2). Across subjects, we observed a significant build-up in the sentence condition ($p < 0.05$), but not in the other three conditions. The sentences build-up effect was individually significant in Subject A1 ($p < 0.01$), Subject A2 ($p < 0.05$), and in the one EOI of Subject A3 ($p < 0.01$).

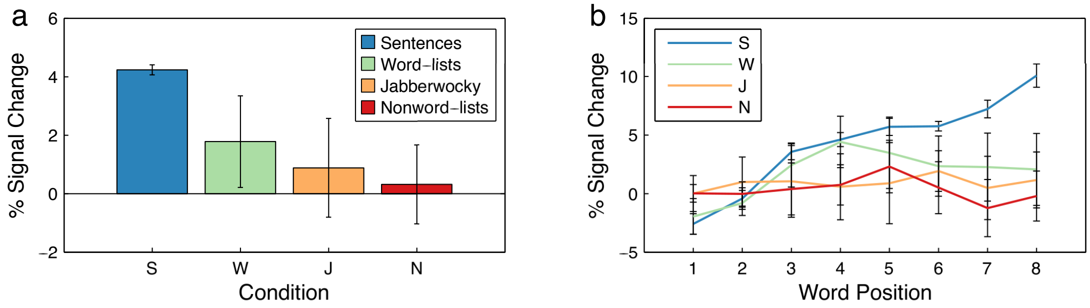

Figure SI-8. a. Gamma magnitude for the auditory conditions averaged across word positions and subjects. EOIs were selected based on the data from the visual trials.
b. Gamma magnitude for the auditory conditions averaged across subjects in each of eight word positions.
Error bars indicate standard errors of the mean over subjects in both panels.

	S - Vis	W - Vis	J - Vis	N - Vis	S - Aud	W - Aud	J - Aud	N - Aud
Subject A1	98.44%	84.38%	79.69%	75.00%	98.44%	93.73%	89.06%	78.13%
Subject A2	98.44%	85.94%	76.56%	75.00%	95.31%	85.94%	82.81%	78.13%
Subject A3	96.88%	93.75%	78.13%	71.88%	92.19%	90.63%	87.50%	81.25%

Table SI-10. Percent accuracy on the behavioral task across all runs of the experiment.

Experimental Procedures

Participants

We recorded electrical activity from intracranial electrodes of three subjects (2 female, ages 19-57) with intractable epilepsy who read or listened to sentences, lists of words, Jabberwocky sentences, and lists of nonwords. These subjects underwent temporary implantation of subdural electrode arrays at Albany Medical College (AMC) to localize the epileptogenic zone(s) and to delineate it from eloquent cortical areas prior to brain resection. All of the subjects gave informed written consent to participate in the study, which was approved by the Institutional Review Board of Albany Medical College.

Materials and procedure

In an event-related design, subjects read or listened to sentences, lists of words, Jabberwocky sentences, and lists of nonwords. The materials for the visual trials were the same as those used in the original experiment and the procedure was identical; the materials for the auditory trials were recorded as described in [1].

Presentation rates of visual trials were consistent across the three new subjects and corresponded to the fast timing version of the experiment described in the main text of this article. Briefly, subjects viewed each word/nonword for 450 ms, then, after each 8-item sequence (lasting for a total of 3,600 ms), a fixation cross was presented for 250 ms followed by the probe item which remained on the screen for 1400 ms and a post-probe fixation cross (250 ms). Each auditory trial contained a single 8-item sequence with mean duration 2,671 ms (SD = 825 ms) and silence added at the end, bringing the total trial duration to 4500 ms. This was followed by 250 ms of fixation, a 1500 ms auditory probe period, and a 250 ms post-probe fixation. Behavioral responses were continually recorded. After each trial, a fixation cross was presented for a variable amount of time, semi-randomly selected from a set of durations (1500, 2000, 2500, or 3000 ms), in order to obtain a low-level baseline for neural activity.

Trials were grouped into runs in order to give participants short breaks throughout the experiment. Each run included four trials per condition per modality and lasted 284 sec. The total amount of inter-trial fixation in each run was 72 sec. All subjects completed sixteen runs of the experiment, for a total of 64 trials per condition per modality.

Data collection and analysis
The implanted electrode grids consisted of platinum-iridium electrodes that were 4mm in diameter (2.3-3 mm exposed) and spaced with an inter-electrode distance of 0.6 or 1 cm. The total numbers of implanted electrodes were 90, 114, and 97 for the three subjects, respectively. Electrodes were implanted on the left hemisphere. Signals were digitized at 1200 Hz. Recordings were synchronized with stimulus presentation and stored using the BCI2000 software platform [6-7]. Upon visual inspection of the recordings, we removed reference electrodes, ground, and electrodes with high noise levels and interictal activity as revealed by independent analyses, which left 89, 108, and 91 electrodes for the three subjects.

Cortical mapping; Extraction of ECoG signal
See Experimental Procedures in the main text.

Selecting electrodes of interest (EOIs)
See Experimental Procedures in the main text. The only difference was that all of the visual data were used for defining the EOIs, and the auditory data were then used to test for generalization of the key patterns to the auditory presentation modality.

i. Examining electrodes that show sensitivity to difficulty

As discussed in the manuscript, the EOIs show a response profile that cannot be explained by difficulty. In fact, the condition that produces the strongest response (the *sentence* condition) is the easiest one. However, one might expect to find some electrodes that are sensitive to difficulty (e.g., [8]). To search for such electrodes, we examined the following contrasts: word-lists > sentences (W>S), Jabberwocky > sentences (J>S), and nonword-lists > sentences (N>S). In the behavioral data, we found substantially lower accuracies in the memory probe task for the Jabberwocky and nonword-list conditions than the sentence condition (10-15% lower), indicating that as expected, Jabberwocky and nonword-list conditions were more difficult than the sentence conditions. In prior work [1], we had also observed lower accuracies in the word-list condition than in the sentence condition, so although these two conditions produced similarly high accuracies in the current set of participants, we included that contrast for completeness.

Figure SI-9 shows the locations of the electrodes (across all six subjects) that showed a significant effect (at $p<0.01$, uncorrected level) for one or both of the target contrasts (J>S, N>S) in the odd runs data (the W>S electrodes did not show replicable responses in the even runs data, as shown below, and were therefore not included in this figure). Table SI-11 shows the information on how many electrodes each subject contributes to each set.

The electrodes are concentrated in the superior temporal regions and in parts of the frontal lobe. The frontal lobe locations are highly consistent with the activation patterns we have previously observed in fMRI for similar contrasts, including the opercular portion of the inferior frontal gyrus and parts of the middle frontal gyrus (e.g., Fig. S1 top panel in [9], for the nonword-lists > sentences contrast). These locations plausibly correspond to the domain-general multiple demand (MD) brain regions [8].

The superior temporal locations are more surprising, as we have not observed responses there in our prior fMRI work. We tend to subvocalize when reading, and constructing a sound-level representation of e.g., unfamiliar nonwords is plausibly more demanding. This may lead to greater responses in some parts of the auditory cortex. However, the absence of similar effects in fMRI is puzzling.

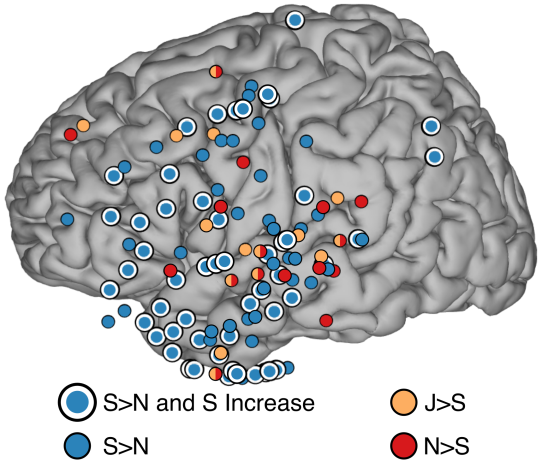

Figure SI-9. A Talairach cortical model showing electrodes across the six participants that showed a significant (at $p<0.01$) J>S effect (orange) or N>S effect (red) in the odd runs data (electrodes colored in red and orange showed both the J>S and the N>S effect), along with the EOIs and the sentences > nonword-list electrodes for reference. The apparent spatial overlap of S>N and N>S/J>S electrodes is in line with prior fMRI work showing that language-selective and domain-general effort-sensitive regions can lie in close proximity to each other, and the high inter-individual variability in the locations of these regions [10].

Having selected some electrodes of interest based on the odd runs data, we examined the robustness of these effects by looking at the responses of these electrodes to the relevant conditions in the even runs data. The responses are shown in Figure SI-10. The J>S and N>S effects appear to be robust showing replicable responses, but the W>S effect does not show a replicable difference in the even runs data, consistent with the fact that the former two contrasts, but not the W>S contrast manifested behaviorally (in the memory probe task performance).

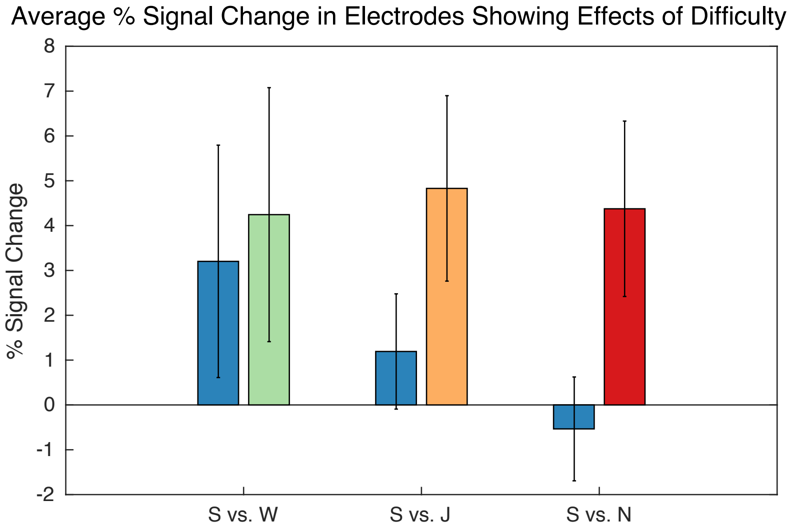

Figure SI-10. Gamma magnitude for the relevant conditions in the even runs data in the electrodes identified based on the odd runs data (i.e., to the W and S conditions in the electrodes that showed a significant W>S effect in the odd runs data, etc.). Two participants contributed W>S electrodes (n=8 electrodes), four participants J>S electrodes (n=15), and five participants contributed N>S electrodes (n=16). Error bars indicate standard errors of the mean over electrodes.

Subject	**W > S**	**J > S**	**N > S**
Subject 1	None	None	1
Subject 2	5 (1 overlap with N>S and J>S) (1 overlap with J>S) (1 overlap with N>S)	4 (1 overlap with W>S and N>S) (1 overlap with W>S) (1 overlap with N>S)	7 (1 overlap with W>S and J>S) (1 overlap with W>S) (1 overlap with J>S)

Subject 3	None	6 (3 overlap with N>S)	3 (3 overlap with J>S)
Subject 4	None	1	3
Subject 5	3 (1 overlap J>S) (1 overlap N>S)	4 (1 overlap W>S) (1 overlap N>S)	2 (1 overlap W>S) (1 overlap J>S)
Subject 6	None	None	None

Table SI-11. Breakdown of electrodes identified in ODD runs exhibiting significant effects of difficulty in the odd runs data.

References:

1. Fedorenko, E., Hsieh, P.J., Nieto-Castañón, A., Whitfield-Gabrieli, S., and Kanwisher, N. (2010). New method for fMRI investigations of language: defining ROIs functionally in individual subjects. J. Neurophysiol. *104*, 1177-1194.

2. Levy, R. (2008). Expectation-based syntactic comprehension. Cognition. *106*, 1127-1177.

3. Smith, N.J., and Levy, R. (2013). The effect of word predictability on reading time is logarithmic. Cognition. *128*, 302-319.

4. Kutas, M. and Federmeier, K.D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. *62*, 621-647.

5. Braze, D., Mencl, W. E., Tabor, W., Pugh, K. R., Constable, R. T., Fulbright, R. K., ... & Shankweiler, D. P. (2011). Unification of sentence processing via ear and eye: an fMRI study. Cortex. *47*(4), 416-431.

6. Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., and Wolpaw, J.R. (2004). BCI2000: A general-purpose brain-computer interface (BCI) system. IEEE Trans. Biomed. Eng. *51,* 1034-1043.

7. Schalk, G. and Mellinger, J. A. (2010). Practical guide to brain-computer interfacing with BCI2000. (London: Springer London).

8. Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends Cogn. Sci. *14* (4), 172–179.

9. Fedorenko, E., Duncan, J. & Kanwisher, N. (2013). Broad domain-generality in focal regions of frontal and parietal cortex. Proc. Natl. Acad. Sci. U. S. A. 110 (41), 16616–16621.

10. Fedorenko, E., Duncan, J. & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. Curr Biol, *22*, 2059-2062.