Neurophysiological dynamics of phrase-structure building during sentence processing

Matthew J. Nelson[a,1,2], Imen El Karoui[b], Kristof Giber[c], Xiaofang Yang[d], Laurent Cohen[b,e], Hilda Koopman[f], Sydney S. Cash[c], Lionel Naccache[b,e], John T. Hale[g], Christophe Pallier[a,h], and Stanislas Dehaene[a,h,i,1]

[a]Cognitive Neuroimaging Unit, Commissariat à l'Energie Atomique (CEA), Institut National de la Santé et de la Recherche Médicale (INSERM) U992, NeuroSpin Center, 91191 Gif-sur-Yvette, France; [b]Institut National de la Santé et de la Recherche Médicale (INSERM) U1127, Centre National de la Recherche Scientifique (CNRS) UMR7225, Université Pierre et Marie Curie (UPMC) Paris 6, Institut du Cerveau et de la Moelle Épinière Research Center (ICM), 75013 Paris, France; [c]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; [d]Stanford Human Intracranial Cognitive Electrophysiology Program (SHICEP), Stanford, CA 94305; [e]AP-HP Groupe hospitalier Pitié-Salpêtrière, 75013 Paris, France; [f]Department of Linguistics, University of California, Los Angeles, CA 90095; [g]Department of Linguistics, Cornell University, Ithaca, NY 14853; [h]Université Paris 11, 91400 Orsay, France; and [i]Collège de France, 75005 Paris, France

Although sentences unfold sequentially, one word at a time, most linguistic theories propose that their underlying syntactic structure involves a tree of nested phrases rather than a linear sequence of words. Whether and how the brain builds such structures, however, remains largely unknown. Here, we used human intracranial recordings and visual word-by-word presentation of sentences and word lists to investigate how left-hemispheric brain activity varies during the formation of phrase structures. In a broad set of language-related areas, comprising multiple superior temporal and inferior frontal sites, high-gamma power increased with each successive word in a sentence but decreased suddenly whenever words could be merged into a phrase. Regression analyses showed that each additional word or multiword phrase contributed a similar amount of additional brain activity, providing evidence for a merge operation that applies equally to linguistic objects of arbitrary complexity. More superficial models of language, based solely on sequential transition probability over lexical and syntactic categories, only captured activity in the posterior middle temporal gyrus. Formal model comparison indicated that the model of multiword phrase construction provided a better fit than probability-based models at most sites in superior temporal and inferior frontal cortices. Activity in those regions was consistent with a neural implementation of a bottom-up or left-corner parser of the incoming language stream. Our results provide initial intracranial evidence for the neurophysiological reality of the merge operation postulated by linguists and suggest that the brain compresses syntactically well-formed sequences of words into a hierarchy of nested phrases.

intracranial | merge | constituent | neurolinguistics | open nodes

Most linguistic theories hold that the proper theoretical description of sentences is not a linear sequence of words, in the way we encounter it during reading or listening, but rather a hierarchical structure of nested phrases (1–4). Whether and how the brain encodes such nested structures during language comprehension, however, remains largely unknown. Brain-imaging studies of syntax have homed in on a narrow set of left-hemisphere areas (5–16), particularly the left superior temporal sulcus (STS) and inferior frontal gyrus (IFG), whose activation correlates with predictors of syntactic complexity (6, 7, 10, 13, 14). In particular, core syntax areas in left IFG and posterior STS (pSTS) show an increasing activation with the number of words that can be integrated into a well-formed phrase (10, 14). Similarly, magneto-encephalography signals show increasing power in beta and theta bands during sentence-structure build-up (17) and a systematic phase locking to phrase structure in the low-frequency domain (18).

These studies leave open the central question of whether and how neural populations in these brain areas create hierarchical phrase structures within each sentence. To address this question, intracranial recordings with more precise joint spatial and temporal resolution may be necessary. The intracranial approach to language processing has led to important advances in phonetics

(19) and single-word processing (20). At the sentence level, a monotonic increase in activity, broadly distributed in the left hemisphere, was found during sentence processing (21), but its relation to sentence-internal structure-building operations remained unexplored. An entrainment of magneto-encephalographic and intracranial responses at the frequency with which phrase structures recurred in short sentences was recently reported (18). However, such frequency-domain measurements are indirect, and the specific time-domain computations leading to such entrainment, and their brain localization, remain unclear. Here, we used direct time-domain analyses of intracranial neurophysiological data to shed light on the neural activity underlying linguistic phrase-structure formation.

Our research tests the hypothesis that during comprehension people "parse" the incoming sequence of words in a sentence into a tree-like structure that captures the part–whole relationships between syntactic constituents. This basic idea has been at the heart of psycholinguistics since the Chomskyan revolution of the 1960s (see, e.g., ref. 22). To understand this sense of "parsing," consider the example "Ten sad students of Bill Gates" shown in Fig. 1*A*. These six words are grouped both semantically
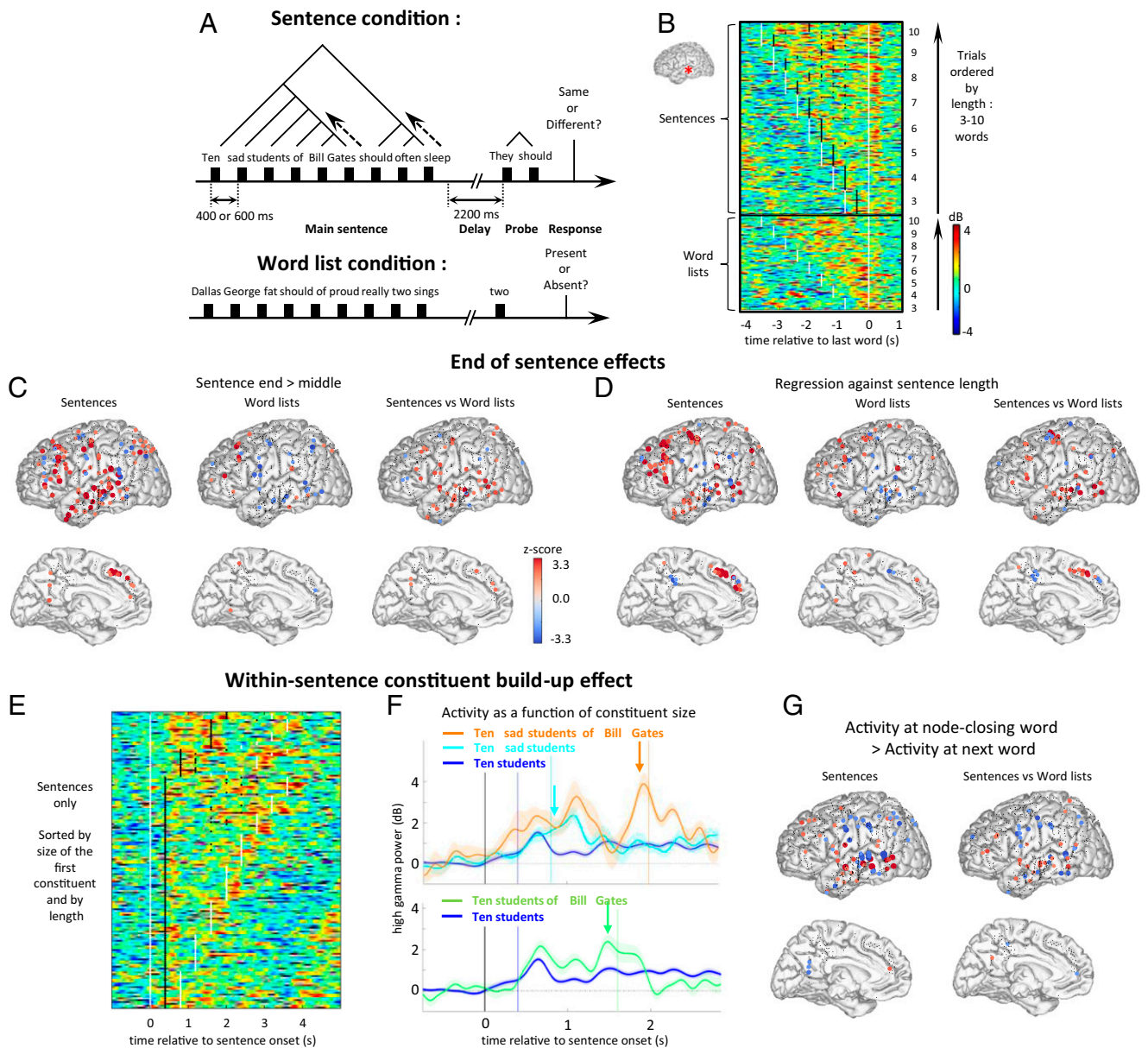
Fig. 1. Experimental setup. (A) Patients saw sentences or word lists of variable length and judged whether a subsequent probe matched. Each word appeared at a fixed location and temporal rate mimicking a moderate, easily comprehensible rate of natural speech or reading (37). (B) High-gamma power recorded from an STS electrode in one subject, aligned to the onset of the last word. Rows in the matrix show individual trials, sorted by condition (sentence or list) and length (number of words). White dots indicate the first and last words of the stimulus. Black dots indicate the onset time of words that allowed a constituent to close. (C and D) Localization of electrodes showing end-of-sentence effects. Each dot represents an electrode (aggregated across subjects), with size and color saturation reflecting the corresponding z score. The z scores below an absolute value threshold of 1.96 are plotted with a black dot. Electrodes located >10 mm from midline are projected onto the lateral surface plot (*Top*); others are projected onto the midline plot (*Bottom*). (C) Contrast of greater activity 200–500 ms after the last word than after other words, either within sentences (*Left*), word lists (*Middle*), or when contrasting sentences with word lists (*Right*). (D) Effect of sentence length on activity evoked 200–500 ms after the final word (same format). (E) Trial-by-trial image of the same aSTS electrode, now aligned to sentence onset and sorted by size of the first constituent. (F) Each trace shows activity averaged across trials with a given size of the first constituent for the same electrode. Colored vertical lines indicate the onset of the corresponding last word of each first constituent. Example sentence beginnings for each trace are shown. Shaded regions indicate the SEM. For illustrative purposes, the time at which a 300-ms sliding window *t* test of activity differences between selected traces exceeded a z-score threshold of 1.96 is indicated with vertical arrows. *Bottom* shows traces where the first constituent is either 2 words (e.g., "Ten students") or 5 words ("Ten students of Bill Gates") long, whereas *Top* shows traces where the first constituent is 2, 3 ("Ten sad students"), or 6 words ("Ten sad students of Bill Gates") long. In *Top*, the light blue and orange arrows mark the difference timepoint between 2- vs. 3-word and 3- vs. 6-word constituents, respectively. (G) Contrast of activity for word Nc minus the activity for word Nc+1, either within sentences (*Left*) or when contrasting sentences with word lists (*Right*).

and syntactically into a single noun phrase, which later serves as a subject for the subsequent verb phrase. We sought to identify the neural correlates of this operation, which we call here "merge" in the sense of Chomsky's Minimalist Program (23). Note, however, that our research does not assume this research program in full but addresses the more general question of whether and how the human brain builds nested phrase structures.

Specifically, we reasoned that a merge operation should occur shortly after the last word of each syntactic constituent (i.e., each phrase). When this occurs, all of the unmerged nodes in the tree

comprising a phrase (which we refer to as "open nodes") should be reduced to a single hierarchically higher node, which becomes available for future merges into more complex phrases. In support of this idea, psycholinguistic studies have identified several behavioral markers of phrasal boundaries (24, 25) and sentence boundaries (26), which are the precise moments when syntactic analyses suggest that merges take place. We hypothesize that areas of the left hemisphere language network might be responsible for this merge operation during sentence comprehension.

How might these areas enact the merge operation? We expected the items available to be merged (open nodes) to be actively maintained in working memory. Populations of neurons coding for the open nodes should therefore have an activation profile that builds up for successive words, dips following each merge, and rises again as new words are presented. Such an activation profile could follow if words and phrases in a sentence are encoded by sparse overlapping vectors of activity over a population of neurons (27, 28). Populations of neurons involved in enacting the merge operation would be expected to show activation at the end of constituents, proportional to the number of nodes being merged. Thus, we searched for systematic increases and decreases in brain activity as a function of the number of words inside phrases and at phrasal boundaries.

We pitted this phrase-structure model against several alternative models of language. Some scholars contest the importance of nested hierarchical structures and propose that language structures can be captured solely based on the transition probabilities between individual words or word groups. N-gram Markov models, for example, describe language based solely on the probabilistic dependency of the current word with a fixed number of preceding words (29). Two variables can be computed from these models that should capture brain activity: entropy, which measures uncertainty before observing the next word, and surprisal, which measures the improbability of observing a specific word. Indeed, these variables have been previously shown to correlate with brain activity (8, 30–33). We therefore conducted an extensive evaluation of their ability to account for the present intracranial data, using various ways of computing transition probabilities, either at the lexical or at the syntactic-category level.

Alternatively, if the brain does in fact implement a parsing algorithm as an integral part of normal comprehension (34), then intracranial data might allow one to narrow down the parsing strategy. Several computational strategies for syntactic parsing have been given a precise formalization in the context of stack-based push-down automata (35). These strategies, called "bottom-up," "top-down," and "left-corner," make different predictions as to the precise number of parser operations and stack length after each word. Taking advantage of the high temporal resolution of our intracranial recordings, which allowed us to measure the activation after each word, we used model comparison to formally evaluate which parsing strategy provided the best account of left-hemisphere brain activity.

Results

We recorded from a total of 721 left-hemispheric electrodes [433 depth electrodes recording stereo EEG and 288 subdural electrodes recording the electro-corticogram (ECOG)] in 12 patients implanted with electrodes as part of their clinical treatment for intractable epilepsy (*SI Appendix*, Fig. S1). Patients read simple sentences, 3–10 words long, presented word-by-word via Rapid Serial Visual Presentation (Fig. 1*A*) in their native language (English or French), at a comfortable rate of 400 or 600 ms per word. Each sentence (e.g., "Bill Gates slept in Paris") was followed after 2.2 s by a shorter elliptic probe, and participants decided whether it matched or did not match the previous sentence (e.g., "he did" versus "they should"). This task was selected because it was intuitive to the patients yet required

understanding the sentence and committing its structure to working memory for a few seconds.

To generate the stimuli, we created a Matlab program rooted in X-bar theory (3, 36) that generated syntactically labeled sentences combining multiple X-bar structures into constituents of variable size, starting from a minimal lexicon (see *SI Appendix* for details). We used the constituent structures created during the generation of the stimuli as a first hypothesis about their internal syntactic trees.

As a control, subjects also performed a working memory task in which they were presented with lists of the same words in random order, then 2.2 s later with a word probe. Participants decided whether or not that word was in the preceding list. This control task had the same perceptual and motor demands as the sentence task but lacked syntax and combinatorial semantics.

We analyzed broadband high-gamma power, a proxy for the overall population activation near a recording site.

End-of-Sentence Effects. We first looked at activity toward the end of the sentence, where a series of merge operations must occur to construct a representation of the entire utterance. Fig. 1*B* shows the trial-by-trial responses from an electrode over the left STS, time-locked to the onset of the last word, separately for sentences and word lists, with stimuli sorted according to the total number of words. A peak in activation systematically followed the last word of the sentence, and that peak became increasingly higher as sentence length increased. Both effects were stronger in the sentence task than in the word list task (sentence end > sentence middle: $P < 0.01$; regression of sentence end activity against sentence length: $P < 0.05$). When all electrode sites were tested, both effects were found to be significant across electrodes and subjects (Fig. 1 *C* and *D* and *SI Appendix*, Table S1) at a subset of sites that largely overlap with the classical fMRI-based localization of the language areas involved in syntactic and semantic aspects of sentence processing (5–16)—that is, the full extent of the STS and neighboring temporal pole (TP), superior temporal and middle temporal sites, and the IFG (primarily pars triangularis, but with scattered activity in orbitalis and opercularis). Additional electrodes with significant effects were found at a cluster of superior precentral and prefrontal sites, previously observed in fMRI (15) and roughly corresponding to the recently reported language-responsive area 55b (37). In summary, both frontal and temporal areas appeared to strongly activate toward sentence ending in proportion to the amount of information being merged. This end-of-sentence effect appears to be a potential neurophysiological correlate of the behavioral finding of a selective slowing of reading time on the last word of a sentence (26).

Within-Sentence Effects of Constituent Closure. We next examined the prediction that merge operations should be observed not only at sentence ending but also within the sentence itself, whenever a constituent can be closed. Fig. 1*E* shows activity in the same STS electrode when trials are sorted according to the size of the initial noun phrase. Averaging across trials (Fig. 1*F*) confirmed a progressive build-up of activity followed by a drop locked to the final merge. For each electrode, we contrasted activity after the appearance of the last word in a constituent (hereafter word *Nc*) with activity after the succeeding word (word *Nc+1*) (Fig. 1*G*). The drop from *Nc* to *Nc+1* in the main task was significant across electrodes and subjects in the anterior STS (aSTS) and pSTS regions (*SI Appendix*, Fig. S2*A*). The effect was also significant across electrodes in the TPJ and across subjects in the IFG pars triangularis (IFGtri).

A similar drop in activation was seen surrounding words that could have potentially been the last word of a constituent (word *pNc*, standing for "potential node close")—for instance, the word "students" in "ten sad students of Bill Gates." Before receiving the word "of," indicating the start of a complement, a parser of

the sentence might transiently conclude that "ten sad students" is a closed constituent (i.e., a fully formed multiword phrase). Indeed, we found a transient decrease in activity following word pNc+1 compared with word pNc that was significant across electrodes and subjects in the IFG pars orbitalis (IFGorb) and pSTS (*SI Appendix*, Fig. S2B), although the IFGorb only approaches significance when correcting for multiple comparisons across regions of interest (ROIs). This effect was also significant across electrodes in the IFGtri and approaching significance across subjects in the aSTS. This finding supports the idea that brain activity reflects the formation of constituent structures by a parser that processes the sentence word-by-word, without advanced knowledge of the overall constituent lengths of each sentence.

Comparison Between Sentences and Word Lists. To control for the possibility that such a drop would occur merely as a function of ordinal word position or due to the identity or grammatical category of the final word, we conducted a "sham" analysis using data from the word-list condition: For each Nc word in the sentence task, we identified a corresponding sham word in the control task, either based on the same word identity or the same ordinal word position (figures and statistics refer to the former; see *SI Appendix*, Fig. S4 for the latter). The drop from Nc to Nc+1 in the main task was significantly larger than in the control task for this electrode ($P < 0.01$). Across all electrodes, these effects were significant in the aSTS and pSTS regions (*SI Appendix*, Fig. S2).

In the word-list task, with high working-memory demands but no possibility of constituent-structure building, trial-by-trial evoked responses in aSTS (Fig. 1B) showed that drops like those following constituent unifying words in the sentence task did not occur. Rather, there was sustained activity in the aSTS up to about the fifth word presentation during the word list task (*SI Appendix*, Fig. S3A). Thus, across electrodes and subjects, the activation following words 2–5 was lower in sentences than in the word list task in the TP, aSTS, pSTS IFGtri, and IFGorb ROIs (*SI Appendix*, Fig. S3B), possibly reflecting a compression resulting from the node-closing operations in sentences. This effect flipped for words 6–10 in TP and aSTS, where activity continued to increase in the sentence condition and became significantly higher in sentences than word lists. Those findings are congruent with a previous report of a systematic increase in intracranial high-gamma activity across successive words in sentences but not word lists (21) and suggest that working memory demands may saturate after five unrelated words.

Effects of the Number of Open Nodes and Node Closings. A parsimonious explanation of the activation profiles in these left temporal regions is that brain activity following each word is a monotonic function of the current number of open nodes at that point in the sentence (i.e., the number of words or phrases that remain to be merged). We therefore measured the correlations of activity with the number of open nodes at the moment when each word was presented (Fig. 2A). In our sample STS electrode (Fig. 2B), following each word presentation, activation increased approximately monotonically with the number of open nodes present for that word before new merging operations take place. At this site, the peak effect of the number of open nodes on high-gamma power was observed ~300 ms after the presentation of the word.

At other sites, visual inspection suggested that high-gamma power following each word was dominated by a transient burst of activity at the moment of merging, going beyond what could be explained by the number of open nodes. For illustration, Fig. 2D shows one such electrode, located in the IFGtri. We regressed the high-gamma activity at this site against the number of nodes closing at each word after normalizing the activity by the number of open nodes present. Both within the sentence and at its very end, activity increased transiently with the numbers of nodes

being closed, again with a peak effect occurring ~300 ms following the critical word's onset (Fig. 2D).

To dissociate these two effects and to examine them at the whole-brain level, we performed a linear multiple regression analysis of the high-gamma power following each word, with the number of open nodes and the number of nodes closing as regressors of interest, and with grammatical category as a covariable of noninterest. For the number of open nodes, coefficients were significantly positive in distributed areas throughout much of the left hemisphere language network, including TP, aSTS, IFG, superior frontal medial and precentral regions, and to a lesser extent the pSTS (Fig. 2C and *SI Appendix*, Table S2). The temporo-parietal junction (TPJ) showed a significant negative effect, suggesting a ramping down of activity in this brain area while the other areas simultaneously ramped up. Altogether, the putative representation of open nodes was mostly confined to classical areas of the sentence-processing network identified with fMRI (14, 15), namely the STS, IFG, dorsal precentral/prefrontal cortex in the vicinity of area 55b (37), and the dorsomedial surface of the frontal lobe. A previous fMRI study suggested that this area may maintain a linguistic buffer (38), which is consistent with a representation of open nodes.

For the number of nodes closing, the most prominent effect was observed in the IFGtri, where it was significant across electrodes and across subjects. IFGorb, pSTS, and TPJ also show significant positive node-closing effects by some measures, whereas aSTS showed a significant negative effect. A sham analysis, consisting in applying those regressors to activity evoked by the same words in the word-list condition, demonstrated the specificity of these findings to the sentence condition (Fig. 2E and *SI Appendix*, Table S2 and Fig. S4).

Impact of Single Words and Multiword Phrases on Brain Activity. The above regressions, using "total number of open nodes" as an independent variable, were motivated by our hypothesis that a single word and a multiword phrase, once merged, contribute the same amount to total brain activity. This hypothesis is in line with the notion of a single merge operation that applies recursively to linguistic objects of arbitrary complexity, from words to phrases, thus accounting for the generative power of language (23). In the present work, multiple regression afforded a test of this hypothesis. We replicated the above multiple regression analysis while separating the total number of open nodes into two distinct variables: one for the number of individual words remaining to be merged, and another for the number of remaining closed constituents comprising several merged words (Fig. 3A).

The results indicated that both variables made independent positive contributions to brain activity at most sites along the STS and IFG (Fig. 3 A–C). Furthermore, their two regression weights were highly correlated across electrodes and with a slope close to 1 (1.23 ± 0.12, 95% confidence interval), thus confirming that words and phrases have similar weights at most sites. At other sites, however, such as the precuneus and inferior parietal cortex, only the effect of the number of closed constituents reached significance. A direct comparison (Fig. 3D and *SI Appendix*, Table S3) indicated that some areas showed a significantly greater sensitivity to constituents than to individual words: the TP, precuneus, IFG, precentral gyrus and inferior parietal lobule (IPL). These findings are congruent with previous indications that those areas may be more specifically involved in sentence or even discourse-level integration (39). Conversely, the pSTS/middle temporal gyrus showed a stronger responsivity to individual words (Fig. 3D and *SI Appendix*, Table S3).

Evaluating the Fit of Models Based on Transition Probabilities. As noted in the Introduction, not all linguists agree that nested phrase structures provide an adequate description of language. Some argue that the appropriate description is sequence-based
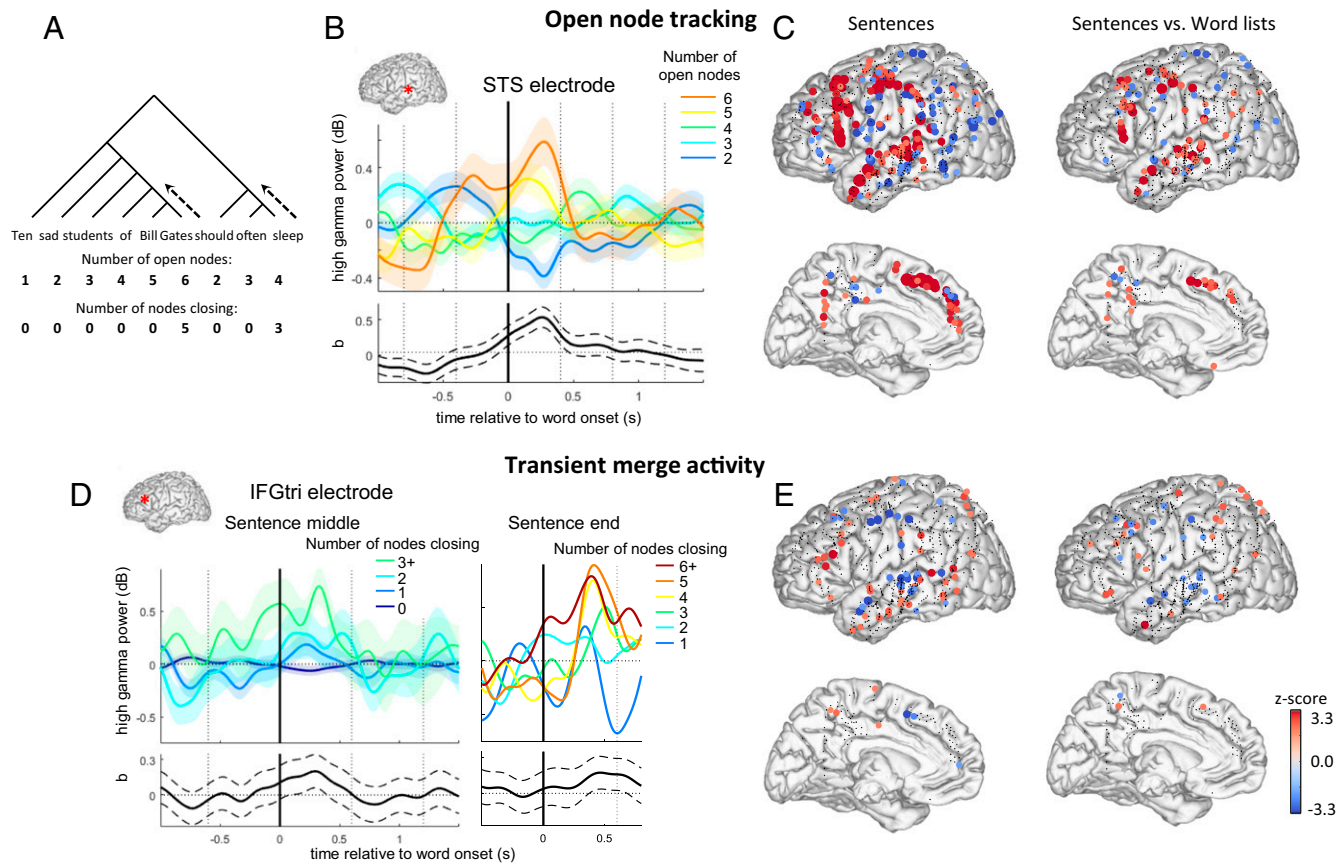
Fig. 2. Tracking of open nodes and nodes closing. (*A*) The number of open nodes and number of nodes closing upon each word are shown below the tree structure diagram for an example sentence. These variables are related by the expression that for word *n* in the sentence: $\text{nOpenNodes}_n = \text{nOpenNodes}_{n-1} - \text{nNodesClosing}_{n-1} + 1$, with boundary conditions that $\text{nOpenNodes}_0 = 0$, $\text{nNodesClosing}_0 = 0$, and $\text{nOpenNodes}_{k+1} = 1$ for a sentence of length k. (*B*) For an example STS electrode, each trace in *Upper* shows the average profile when the corresponding number of open nodes is present upon the presentation of the word at time 0. *Lower* shows the temporal profile of the ANCOVA coefficient evaluating the effect of the number of open nodes on high-gamma power (i.e., the linear relationship from cool to warm colors above) with its 95% confidence interval in dashed lines. To avoid effects from the startup of the sentence from dominating the plot, the data were normalized by word position, and the first two words of the sentence were discarded. (*D*) For an IFG electrode, each trace in *Upper* shows the average profile when the corresponding number of nodes closing upon the presentation of the word at time 0, with the entire trace normalized by the number of open nodes for the word at time 0. *Left* plots show activity during the sentence. *Right* plots show the same for the last word of the sentence only. There were fewer words with a large number of node closings during the sentence than at sentence ending, which necessitated the different groupings between the two plots. The plot below shows the temporal profile of the ANCOVA coefficient measuring the impact of node closings on high-gamma power. (*C* and *E*) Multiple regression model simultaneously modeling the effects of number of open nodes (*C*) and number of nodes being closed (*E*). The z scores for the corresponding regressors are shown for the sentence task (*Left*) and the interaction of the regressors across the sentence and word list tasks (*Right*), where positive indicates a larger coefficient in the sentence task. *SI Appendix*, Table S2 shows the result of significance tests across electrodes and across subjects for each ROI.

and probabilistic (29): Language users would merely rely on probabilistic knowledge about a forthcoming word or syntactic category, given the sequence of preceding words. We therefore performed a systematic comparison of which model provided the better fit to our intracranial data.

We tested four successive models that eschewed the use of hierarchical phrase structure and relied solely on transition probabilities either between specific words or between syntactic categories, using probability estimates derived either from our stimuli or from a larger English corpus, the Google Books Ngrams (Fig. 4 and *SI Appendix*, Figs. S5–S7). The best fitting of these models is shown in Fig. 4. For this model, we applied a first-order Markov model (n-gram model) of the sequential transition probabilities between words in our stimuli, ignoring any potential phrase structure in the sentences. To estimate the transition probabilities, we used a set of 40,000 sentences generated by the same sentence-generating program used to create our stimuli. We used these probabilities to compute both the entropy of the subsequent word and the surprisal of the word just received for every word presented to the patients. These values

were then included in multiple regression analyses, either instead of or in addition to parameters for the number of open nodes and number of node closings (Fig. 4 *A* and *B* and Fig. 4*C*, respectively).

We found a negative effect of entropy (indicating that brain activity increased when certainty of the next word increased) in a posterior temporal region located in the vicinity of pSTS and/or posterior middle temporal gyrus, overlapping with our predefined ROIs for TPJ and pSTS. In these two ROIs, the data were significantly better fit by the n-gram model than with the open-nodes phrase-structure model (*SI Appendix*, Table S4), although this improvement only approaches significance after correction for multiple comparisons across ROIs. We also observed a positive effect of surprisal in the same region, indicating a greater neural activity to unpredicted words. This effect became particularly pronounced in an alternate model when data from the Google Books Ngram corpus (40) was used to determine the transition probabilities (*SI Appendix*, Fig. S5). These observations fit with prior observations of predictive-coding effects in language processing, including single-word semantic
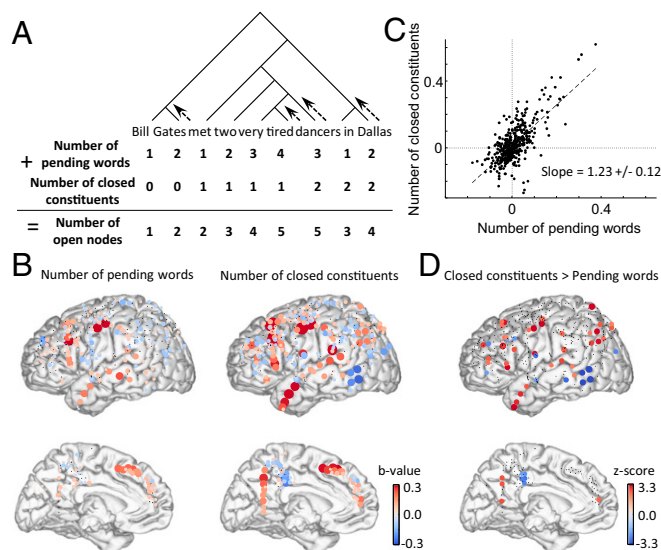
Fig. 3. Effects of number of closed constituents and number of pending words. (*A*) Pictorial description of the evolution of the number of pending words and number of closed constituents for an example sentence. (*B*) The corresponding regression weights (b values) of these two parameters for subjects that received 3–10-word stimuli, using an absolute value threshold of 0.1 dB. (*C*) Scatterplot of the b values for number of closed constituents and number of pending words at each site (each point represents an electrode). (*D*) The z scores for the differences of the coefficients for each electrode, where a positive value indicates a larger coefficient for the number of closed constituents. *SI Appendix*, Table S3 shows the result of significance tests across electrodes and across subjects for each ROI.

priming, N400, and repetition suppression effects in posterior temporal regions (41–45). Other models we tested used the transitions between parts of speech rather than individual words (*SI Appendix*, Fig. S6) and used all of the preceding words in a sentence to compute each transition probability (*SI Appendix*, Fig. S7). These models gave similar but less significant results.

Crucially, however, elsewhere in the language network including the aSTS, TP, IFGtri, and IFGorb areas, activity was systematically fit significantly better by the open nodes model than by any of the transition-probability models (Fig. 4 and *SI Appendix*, Figs. S5–S7 and Table S4). Moreover, the presence of entropy and surprisal as cofactors in a multiple regression did not weaken the coefficients for the number of open nodes or number of nodes closing in any brain area (Fig. 4 and *SI Appendix*, Figs. S5–S7 and Table S4). In the pSTS and TPJ regions that showed a better fit for the lexical bigram model derived from the stimuli, the inclusion of these parameters even resulted in a stronger effect of the number of nodes closing (compare *SI Appendix*, Table S4 to *SI Appendix*, Table S2). The presence of transition-probability effects thus does not appear to detract from the explanatory power of the open-nodes model, and the majority of IFG and STS sites appear to be driven by phrase structure rather than by transition probability.

In conclusion, probabilistic measures of entropy and surprisal that eschewed phrase structure did affect brain activity, in agreement with previous reports (8, 30–33), but they did so primarily in a posterior middle temporal region distinct from the classical brain regions involved in the construction of syntactic and semantic relationships. Furthermore, the effect of probabilistic variables was largely orthogonal to the open-nodes and nodes-closing variables that captured phrase structure.

Evaluating the Fit of Syntactic Parsers. Our findings outline a model whereby, at each step during sentence processing, activation in the

language network reflects the number of words or phrases waiting to be merged as well as the complexity of transient merging operations. However, number of words can only serve as an approximate proxy for syntactic complexity. In the final analysis, brain activation would be expected to reflect the number and the difficulty of the parsing operations used to form syntactic structures, in addition to or even instead of the ultimate complexity of those structures.

We therefore examined whether our intracranial data could be accounted for by an effective syntactic parser. Because a great diversity of parsing strategies have been proposed, we focused on parsers introduced in the context of push-down automata with a stack (35) and contrasted three parsing strategies that all led to the same tree representations: bottom-up parsing, where the rules of a grammar are applied whenever called for by an incoming word; top-down parsing, where a maximum number of rules are applied in advance of each word; and left-corner parsing, which combines both ideas (35). We applied these strategies to the grammar that generated the stimulus sentences and obtained the length of the stack when each given word was received (including that word) and the number of parsing operations that were initiated by the arrival of this word (and before the next one). Those variables were entered into the general linear model instead of the above open-nodes and nodes-closing variables.

Both bottom-up and left-corner parsing provided a good fit of the brain-activity data, at left superior temporal, inferior frontal, dorsal frontal, and midline frontal sites, similar to those captured
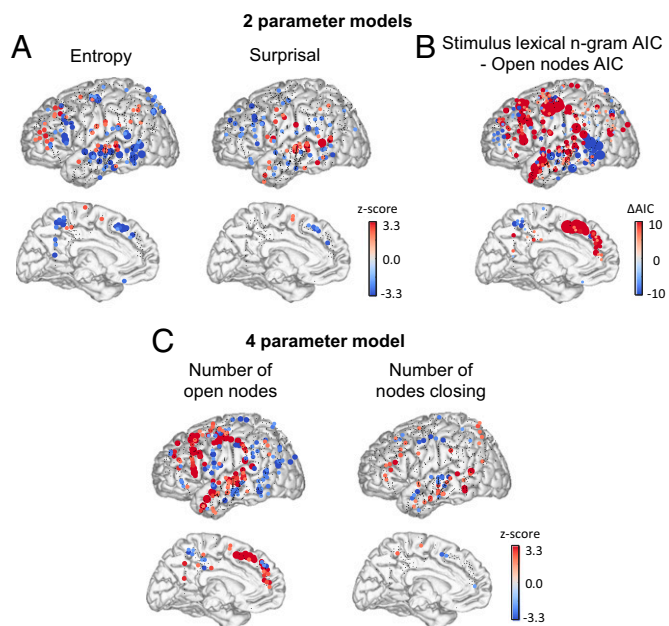
Fig. 4. Stimulus n-gram lexical transition probability model. (*A*) Stimulus n-gram lexical entropy and surprisal coefficient z scores. Plots show the z score for each electrode in the dataset of the coefficient for entropy (*Left*) and surprisal (*Right*) resulting from a model including both parameters in a multiple regression along with other control factors. (*B*) Comparison of model fit to the open-nodes model. Plots show the AIC differences between the two-parameter stimulus lexical n-gram model and the open-nodes model shown in Fig. 2 *C* and *E*. Lower AIC values indicate a better model fit, and therefore, red dots indicate a better fit of the open-nodes model at that electrode site. (*C*) Representation of the number of open nodes and the number of nodes closing in a four-parameter model also incorporating lexical surprisal and entropy as additional control variables. The strength and significance of the effects of open nodes are not much affected by the addition of the stimulus n-gram lexical entropy and surprisal values. *SI Appendix*, Table S5 shows the results of significance tests across electrodes and across subjects for each ROI.

A

Bottom-up

Stack	Operations
[S]	shift John
John [S]	reduce by ProperN➡John
ProperN [S]	reduce by NP➡ProperN
NP [S]	shift loves
loves NP [S]	reduce by V➡loves
V NP [S]	shift Mary
Mary V NP [S]	reduce by ProperN➡Mary
ProperN V NP [S]	reduce by NP➡ProperN
NP V NP [S]	reduce by VP➡V NP
VP NP [S]	reduce by S➡NP VP

Left-corner

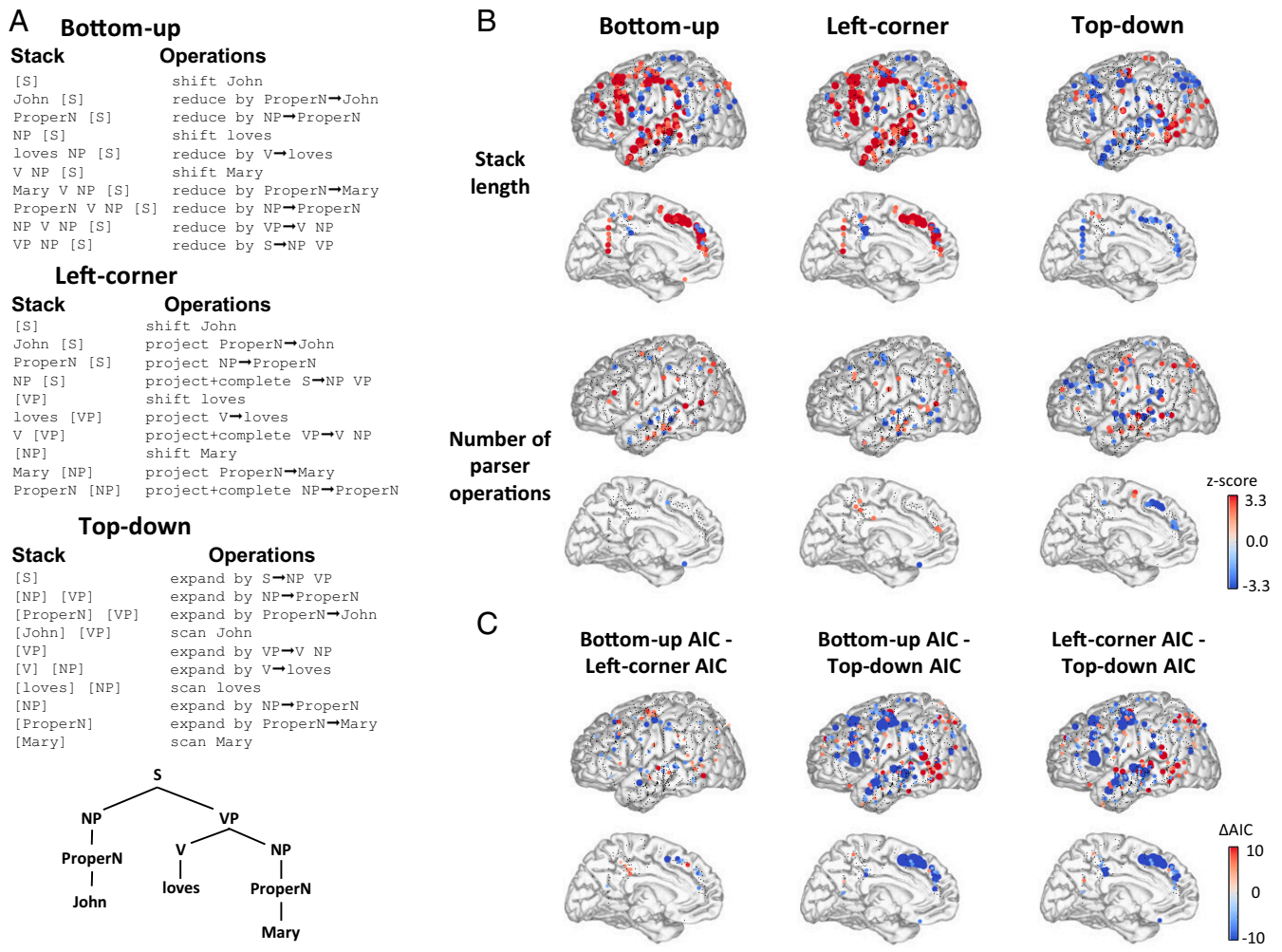| Stack	Operations
[S]	shift John
John [S]	project ProperN➡John
ProperN [S]	project NP➡ProperN
NP [S]	project+complete S➡NP VP
[VP]	shift loves
loves [VP]	project V➡loves
V [VP]	project+complete VP➡V NP
[NP]	shift Mary
Mary [NP]	project ProperN➡Mary
ProperN [NP]	project+complete NP➡ProperN

Top-down

Stack	Operations
[S]	expand by S➡NP VP
[NP] [VP]	expand by NP➡ProperN
[ProperN] [VP]	expand by ProperN➡John
[John] [VP]	scan John
[VP]	expand by VP➡V NP
[V] [NP]	expand by V➡loves
[loves] [NP]	scan loves
[NP]	expand by NP➡ProperN
[ProperN]	expand by ProperN➡Mary
[Mary]	scan Mary

B

Bottom-up   Left-corner   Top-down

Stack length

Number of parser operations

z-score
3.3
0.0
-3.3

C

Bottom-up AIC - Left-corner AIC   Bottom-up AIC - Top-down AIC   Left-corner AIC - Top-down AIC

ΔAIC
10
0
-10

Fig. 5. A test of parsing strategy models. (*A*) Description of the actions taken by a push-down automaton to parse the simple sentence "John loves Mary" using different parsing strategies. NP, noun phrase; ProperN, proper noun; S, sentence; V, verb; VP, verb phrase. (*B*) Coefficients for parsing variables. Each brain shows the distribution of *z* scores associated with a particular coefficient and parsing strategy resulting from a multiple regression analysis applied to sentence data. *Left*, *Middle*, and *Right* show results for bottom-up, left-corner, and top-down parsing strategies, respectively. *Top* and *Bottom* show coefficients for the stack length and number of parser operations, respectively. (*C*) Comparison of model fits between parsing strategies. Each brain shows the distribution of differences in AIC between the pair of models indicated. AIC differences below an absolute value threshold of 4 are plotted with a black dot. A lower AIC indicates a better fit; thus, a more negative difference indicates a better fit for the first of the models being compared (i.e., "Bottom-up" for the comparison "Bottom-up AIC - Left-corner AIC"). *SI Appendix*, Table S6 shows the results of significance tests across electrodes and across subjects for each ROI.

by the open-nodes and node-closing variables (Fig. 5 and *SI Appendix*, Tables S5 and S6). Model comparison supported bottom-up and left-corner parsing as significantly superior to top-down parsing in fitting activation in most regions in this left-hemisphere language network. In the IFGtri region, bottom-up parsing significantly fit the data better across electrodes, owing to a stronger coefficient for the number of parser operations that was significant across electrodes and subjects (only when considering *P* values uncorrected for multiple comparisons across ROIs).

Those findings support bottom-up and/or left-corner parsing as tentative models of how human subjects process the simple sentence structures used here, with some evidence in favor of bottom-up over left-corner parsing. Indeed, the open-node model that we proposed here, where phrase structures are closed at the moment when the last word of a phrase is received, closely parallels the operation of a bottom-up parser. To quantify this similarity, we computed the correlation matrices of the independent variables used in the different models of this study (*SI Appendix*, Table S7). High correlations were found between the bottom-up parsing variables and our node-closing phrase-structure model

($r = 0.97$ between the number of open nodes and stack length; $r = 0.89$ between bottom-up number of parser operations and the number of nodes closing). Moreover, unlike the nonhierarchical probabilistic metrics that we tested (Fig. 4 and *SI Appendix*, Figs. S5–S7 and Table S4), adding the parsing variables to a multiple-regression model including the open-nodes and nodes-closing variables led to a notable decrease in the strength of the node-related coefficients (*SI Appendix*, Fig. S8, compare *SI Appendix*, Table S6 to *SI Appendix*, Table S2). The bottom-up parsing model should therefore not be construed as an alternative to the open-nodes phrase-structure model but rather as an operational description of how constituent structures can be parsed.

Discussion

We analyzed the high-gamma activity evoked at multiple sites of the left hemisphere while French- and English-speaking adults read sentences word-by-word. In a well-defined subset of electrodes, primarily located in superior temporal and inferior frontal cortices, brain activity increased with each successive word but decreased whenever several previous words could be

merged into a syntactic phrase. At this time, an additional burst of activity was seen, primarily in IFGtri, and then the activation dropped, seemingly reflecting a "compression" of the merged words into a single unified phrase. Activation did not drop to zero, however, but remained proportional to the number of nodes that were left to be merged, with a weight that suggested that, in most brain regions, individual words and multiword phrases impose a similar cost on brain activity.

Those results argue in favor of the neurophysiological reality of the linguistic notion of phrase structure. Together with previous linguistic considerations such as syntactic ambiguity, movement, and substitutability (1–4), the present neurophysiological recordings strongly motivate the view of language as being hierarchically structured and make it increasingly implausible to maintain the opposite nonhierarchical view (e.g., refs. 29, 46). With the high temporal resolution afforded by intracranial recordings, phrase structure indeed appears as a major determinant of the dynamic profile of brain activity in language areas. A recent study monitored high-gamma intracranial signals during sentence processing and, after averaging across a variety of sentence structures, concluded that high-gamma activity increases monotonically with the number of words in a sentence (21). The present results suggest that this is only true on average: When time-locking on the onset and offset of phrase structures, we found that this overall increase, in fact, was frequently punctuated by sudden decreases in activity that occur at phrase boundaries.

A quantitative fMRI study previously showed that the average brain oxygen level-dependent (BOLD) signal evoked by an entire series of words increases approximately logarithmically, rather than linearly, with the number of words that can be grouped into constituents (14). The present findings can explain this observation: We now see that a sublinear increase arises from the fact that brain activity does not merely increase every time a new word is added to the sentence but also transiently decreases whenever a phrase-building operation compresses several words into a single node. On average, therefore, the total activity evoked by a sentence is a sublinear and approximately logarithmic function of the number of words it contains (Fig. 6). By encoding nested constituent structures, the brain achieves a compression of the information in sentences and thus avoids the linear saturation that would otherwise occur with an unrelated list of words. The concept of sentence compression can also explain the classical behavioral finding that memory for sentences is better than for random word lists (47), likely reflecting the savings afforded by merging operations. A similar compression effect, called "chunking," has been described when lists of numbers are committed to working memory (48). Similarly, memory for spatial sequences has recently been shown to involve an internal compression based on nested geometrical regularities, with the rate of errors being proportional to "minimal description length"—that is, the size of the internal representation after compression (49). Tree structures and tree-based compression may therefore characterize many domains of human information processing (50, 51).

The present localization of linguistic phrase structures is broadly consistent with prior fMRI studies, which indicate that an array of regions along the STS, as well in the IFG, systematically activate during sentence processing compared with word lists, with a level of activation that increases with the size of linguistic trees (5, 10, 14) and the complexity of the syntactic operations required (6–9, 11–13). A previous fMRI study (52) showed that bilinguals exhibit different neural responses when an incoming auditory stream switches between languages either within a phrase or at a phrase boundary, thus supporting the neural reality of phrase structure. We also observed activations in superior precentral and prefrontal cortex, close to a recently reported language-responsive region called "area 55b" (37), as well as in mesial prefrontal and precuneus areas. All of these
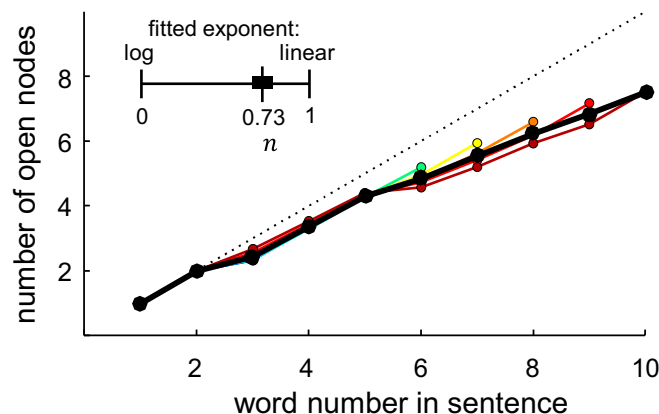

Fig. 6. The average number of open nodes at each word position by sentence length. Lines ranging from cool to hot colors indicate increasing sentence lengths, whereas the solid black line shows the average across sentence lengths. The dotted line shows the unity line (y = x). The nonlinear fit of the log-to-linear parameter is shown above the plot, with a black bar indicating its 95% confidence interval. The number of open nodes grew with an average slope of 0.636 nodes per word. This growth was significantly sublinear, although closer to linear than logarithmic, with a log-to-linear parameter value of 0.73 ± 0.11. If neural activity in the same areas is proportional to the number of open nodes, as the present results suggest, this sublinear effect in the stimuli could explain some of the logarithmic effect observed in Pallier et al. (14).

regions have been previously reported in fMRI studies of sentence processing (7, 15). Consistent with those studies, high-gamma activations to phrase structure were confined to language-related regions and were largely or totally absent from other regions such as dorsolateral prefrontal cortex or dorsal parietal cortex, which have not been reported to be associated with sentence processing in fMRI, although we had good electrode coverage in those regions (see Fig. 2). Still, the distribution of phrase-structure effects was broad, perhaps due to the fact that in our stimuli, syntactic and semantic complexity were confounded. Previous fMRI experiments suggest that, when semantic complexity is minimized by using delexicalized "Jabberwocky" stimuli, fMRI activation becomes restricted to a narrow core-syntax network involving pSTS and IFG (10, 14). In the future, it will be important to examine whether the intracranial signals in those regions continue to respond to Jabberwocky stimuli.

A recent intracranial study revealed an increase in theta power (~5 Hz) in the hippocampus and surrounding cortex associated with semantic predictability during sentence comprehension (53). We found no systematic effect of phrase-structure building in mesial temporal regions in the present data (*SI Appendix*, Fig. S9). Although this negative finding must be considered with caution, it tentatively suggests that the role of those regions may be confined to the storing of semantic knowledge.

We also tested a variety of alternative models sharing the notion that probabilistic predictability, rather than phrase structure, is the driving factor underlying linguistic computations. Those models had a significant predictive value but only in the vicinity of the posterior middle temporal gyrus, whereas activity in other regions continued to be driven primarily by phrase structure. Previous work has repeatedly stressed the importance of predictability in language processing and has indeed pinpointed the posterior middle temporal region as a dominant site for effects of lexico-semantic predictions (8, 30–33, 45). The present work does not question the notion that predictability effects play a major role in language processing, but it does indicate that brain activity in language areas cannot be reduced to such probabilistic measures. Predictability effects, when they occur, are probably driven by nested phrase structures rather than by mere sequential transition

probabilities (8). The model comparison results that we obtained suggest that, in the future, intracranial signals might be used to adjudicate between competing theories of syntactic phrase structures and parsing mechanisms.

Our data provide a first glimpse into the neural mechanisms of syntactic structure building in the brain. Theoretical models of the broadband high-gamma power signal that we measured here suggest that it increases linearly with the number of temporally uncorrelated inputs over a population of neurons near the recording site (54, 55). Its increase with the number of open nodes therefore suggests a recruitment of increasingly larger pools of neurons as syntactic complexity increases. One possible hypothesis bridging from the neural level to the linguistic level is that each open node is represented by a sparse neural vector code and that the vectors for distinct words or phrases are linearly superimposed within the same neural population (27, 28). Under these two assumptions, the broadband high-gamma power would increase monotonically with the number of open nodes present in the sentence, as observed here.

In addition, the present work provides a strong constraint about the neuronal implementation of the merge operation itself: Each merge is reflected by a sudden decrease of high-gamma activity in language areas, related to the changing number of open nodes, accompanied by a transient increase at some IFG and pSTS sites. The decrease may correspond to a recoding of the merged nodes into a new neuronal population vector, orthogonal to the preceding ones, and normalized to the same sparse level of activity as a single word. There are several published proposals as to how such an operation may be implemented (56–58), but the present data cannot adjudicate between them. Higher resolution intracranial recordings (19) or fMRI (59) will be needed to resolve the putative vectors of neural activity that may underlie syntactic operations.

Materials and Methods

Patients. We recorded intracranial cortical activity in the left hemisphere language network from 12 patients performing a language task in their native language (English or French) while implanted with electrodes while awaiting resective surgery as part of their clinical treatment for intractable epilepsy. Written informed consent was obtained from all patients before their participation. Experiments presented here were approved by the corresponding review boards at each institution (the Stanford Institutional Review Board, the Partners Healthcare Institutional Review Board, and the Comité Consultatif de Protection des Personnes for the Pitié-Salpétrière Hospital). Patient details can be seen in *SI Appendix*, Fig. S1.

Recordings. Intracranial recordings were made from ECoG electrodes implanted at the cortical surface at Stanford Medical Center (SMC) and from depth electrodes implanted beneath the cortical surface at Massachusetts General Hospital (MGH) and l'Hopital de la Pitié-Salpêtrière (PS). Electrode positions were localized from a postoperative anatomical MRI normalized to the template MNI brain. To eliminate potential effects from common-mode noise in the online reference and volume conduction duplicating effects in nearby electrodes, recordings were rereferenced offline using a bipolar montage.

Experimental Procedures.
Tasks. In the main sentence task, patients were asked to read and comprehend sentences presented one word at a time using Rapid Serial Visual Presentation. On each trial, the patients were presented with a sentence of variable length (up to 10 words), followed after a delay of 2.2 s by a shorter probe sentence (2–5 words). On 75% of trials, this probe was an ellipsis of the preceding sentence, with the same meaning (e.g., "Bill Gates slept in Paris" followed by "he did"; or "he slept there"; etc.). On 25% of trials, it was an unrelated ellipsis (e.g., "they should"). The subject's task was to press one key for "same meaning" and another key for "different meaning." This task was selected because it required memorizing the entire structure of the target sentence, while being judged natural and easy to perform. Patients also performed a word-list task in which they were presented with a list of words in random order, followed by a single-word probe, and asked to identify whether or not the probe was in the preceding list.

In both conditions, words were presented one at a time at a fixed location on screen (discouraging eye movements). The temporal rate was fixed at 400 ms (four patients) or 600 ms (eight patients) per word, adapted to individual patients' subjective comprehension.

Sentences. We devised a program in Matlab to automatically generate syntactically labeled sentences using a small set of words and syntactic constructions (see Fig. 1*A* and Fig. 3*A* for two example sentences). The program relied on a simplified version of X-bar theory, a classical theory of the organization of constituent structures (36). Starting from an initial clause node, the program applied a series of substitution rules to generate a deep structure for the binary-branching tree representing the syntactic structure of a possible sentence, and a further program generated the sentence's surface structure from this abstract specification. Final sentence length was controlled by selecting an equal number of sentences of each length within the permitted range of sentence lengths (3–10 words). See *SI Appendix, SI Materials and Methods* for more details about the sentences.

For each main sentence, the software also generated automatically a shortened probe sentence by applying up to three randomly selected substitutions, including pronoun substitutions, adjective ellipsis, verb ellipsis (e.g., "he did"), and deictic substitution of location adjuncts (e.g., "they work *there*"). Twenty-five percent of randomly selected trials had their associated shortened probe sentences shuffled between sentences to give nonmatching probes.

Analyses.
Broadband high-gamma power. We analyzed the broadband high-gamma power, which is broadly accepted in the field as a reflection of the average activation and firing rates of the local neuronal population around a recording site (55, 60). Power at each time point was expressed in decibels relative to the overall mean power for the entire experiment. For plotting, we convolved the resulting high-gamma power with a Gaussian kernel with a SD of 75 ms.

ROI analyses. Results were analyzed within ROIs using the regions reported in Pallier et al. (14): TP, aSTS, pSTS, TPJ, IFGtri, IFGorb, superior frontal gyrus (FroSupMed), and a portion of the precentral gyrus (PreCentral). To compensate in part for interindividual variability, electrodes within 22 mm of the group ROI surface were included as part of each ROI, except for the aSTS ROI, which included electrodes within 22 mm of the ROI center. For the analyses summarized in *SI Appendix*, Table S3, ROIs for the precuneus (PreCun) and IPL were added as these areas were anticipated to respond more strongly to more complex constituents. For these areas, we included electrodes within 18 mm of the ROI centers obtained from a different study (61).

General linear model. To evaluate the simultaneous effects of partially correlated variables, we applied a general linear model (multiple regression) using the high-gamma power averaged over the time window of 200–500 ms following each word presentation as the dependent variable. For each model, additional factors of noninterest were introduced for grammatical category and baseline neural activity. Grammatical category was broken down into three classes: closed-class words (determiners and prepositions), open-class words (nouns, verbs, adjectives, and adverbs), and words not readily classified as either open or closed class (numbers and adverbs of degree). We used for the baseline value the average high-gamma power in a 1-s interval before the onset of the first word of each main sentence.

Transition probability models. We used two sources to estimate transition probabilities: either a large set of 40,000 sentences generated by the same program used to create the stimuli, or natural-language data downloaded from the Google Books Ngram Viewer, Version 2 (40). Furthermore, to go beyond bigram models, we developed a model considering the sentence as a string of parts of speech and using the entire sentential prefix to predict the following part of speech in the stimuli (see *SI Appendix, SI Materials and Methods* for details).

Parsing models. We applied three different parsing strategies (bottom-up, left-corner, and top-down) to the stimulus sentences, using the same grammatical rules used to generate the sentences. The outcome was a specification of the order with which each automaton constructed the syntactic tree of a sentence, as shown in Fig. 5*A*. This order was then used to derive, in an automated fashion, both the length of the stack at the moment when a given word was received (including that word) and the number of parsing operations that were initiated by the arrival of this word (and before the next one). These two variables were then entered as covariates of interest in a series of multiple regression analyses.

Statistical analysis. We performed two complementary tests of the z scores of coefficients from the above multiple regression analyses for all electrodes located within a given ROI. We first tested for significance across electrodes using Stouffer's z-score method (62), which tests the significance of the z-score sum across hypothesis tests with an assumption of independence between

tests. We then tested for significance across subjects using a randomization/simulation procedure (see *SI Appendix, SI Materials and Methods*).

For model comparison, we contrasted the Akaike information criterion (AIC) for each model. All models compared had the same number of parameters. To test for a significant AIC difference across electrodes within an ROI, we performed a signed rank test of the median AIC differences across electrodes. To test for a significant AIC difference across subjects within an ROI, we performed a cross-subject permutation test (see *SI Appendix, SI Materials and Methods*). We also compared the overall positivity or negativity of coefficients across electrodes and subjects for models being compared. Across electrodes, we used the two-way version of Stouffer's *z*-score method (62). Across subjects, we used a permutation test to test the mean of mean *z*-score differences across subjects, following the same procedure used to test AIC differences.

1. Berwick RC, Friederici AD, Chomsky N, Bolhuis JJ (2013) Evolution, brain, and the nature of language. *Trends Cogn Sci* 17:89–98.
2. Chomsky N (1957) *Syntactic Structures* (Mouton de Gruyter, Berlin).
3. Sportiche D, Koopman H, Stabler E (2013) *An Introduction to Syntactic Analysis and Theory* (John Wiley & Sons, Hoboken, NJ).
4. Haegeman L (2005) *Thinking Syntactically: A Guide to Argumentation and Analysis* (Wiley–Blackwell, Malden, MA).
5. 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* 23:1859–1873.
6. Ben-Shachar M, Palti D, Grodzinsky Y (2004) Neural correlates of syntactic movement: Converging evidence from two fMRI experiments. *Neuroimage* 21:1320–1336.
7. Brennan J, et al. (2012) Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain Lang* 120:163–173.
8. Brennan JR, Stabler EP, Van Wagenen SE, Luh W-M, Hale JT (2016) Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain Lang* 157-158:81–94.
9. Fedorenko E, Thompson-Schill SL (2014) Reworking the language network. *Trends Cogn Sci* 18:120–126.
10. Goucha T, Friederici AD (2015) The language skeleton after dissecting meaning: A functional segregation within Broca's Area. *Neuroimage* 114:294–302.
11. Hagoort P (2005) On Broca, brain, and binding: A new framework. *Trends Cogn Sci* 9:416–423.
12. Longe O, Randall B, Stamatakis EA, Tyler LK (2007) Grammatical categories in the brain: The role of morphological structure. *Cereb Cortex* 17:1812–1820.
13. Musso M, et al. (2003) Broca's area and the language instinct. *Nat Neurosci* 6:774–781.
14. Pallier C, Devauchelle A-D, Dehaene S (2011) Cortical representation of the constituent structure of sentences. *Proc Natl Acad Sci USA* 108:2522–2527.
15. Jobard G, Vigneau M, Mazoyer B, Tzourio-Mazoyer N (2007) Impact of modality and linguistic complexity during reading and listening tasks. *Neuroimage* 34:784–800.
16. Moro A, et al. (2001) Syntax and the brain: Disentangling grammar by selective anomalies. *Neuroimage* 13:110–118.
17. Bastiaansen M, Magyari L, Hagoort P (2010) Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *J Cogn Neurosci* 22:1333–1347.
18. Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci* 19:158–164.
19. Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. *Science* 343:1006–1010.
20. Sahin NT, Pinker S, Cash SS, Schomer D, Halgren E (2009) Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* 326:445–449.
21. Fedorenko E, et al. (2016) Neural correlate of the construction of sentence meaning. *Proc Natl Acad Sci USA* 113:E6256–E6262.
22. Fodor JA, Bever TG, Garrett MF (1974) *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar* (McGraw-Hill, New York).
23. Chomsky N (2013) Problems of projection. *Lingua* 130:33–49.
24. Bever TG, Lackner JR, Kirk R (1969) The underlying structures of sentences are the primary units of immediate speech processing. *Percept Psychophys* 5:225–234.
25. Abrams K, Bever TG (1969) Syntactic structure modifies attention during speech perception and recognition. *Q J Exp Psychol* 21:280–290.
26. Warren T, White SJ, Reichle ED (2009) Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition* 111:132–137.
27. Smolensky P, Legendre G (2006) *The Harmonic Mind: Cognitive Architecture* (MIT Press, Cambridge, MA).
28. Stewart TC, Bekolay T, Eliasmith C (2011) Neural representations of compositional structures: Representing and manipulating vector spaces with spiking neurons. *Connect Sci* 23:145–153.
29. Frank SL, Bod R, Christiansen MH (2012) How hierarchical is language use? *Proc R Soc Lond B Biol Sci* 279:4522–4531.
30. Frank SL, Otten LJ, Galli G, Vigliocco G (2015) The ERP response to the amount of information conveyed by words in sentences. *Brain Lang* 140:1–11.
31. Hale J (2016) Information-theoretical complexity metrics. *Lang Linguist Compass* 10:397–412.
32. Henderson JM, Choi W, Lowder MW, Ferreira F (2016) Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in reading. *Neuroimage* 132:293–300.
33. Willems RM, Frank SL, Nijhof AD, Hagoort P, van den Bosch A (2016) Prediction during natural language comprehension. *Cereb Cortex* 26:2506–2516.
34. Johnson-Laird P (1989) *The Computer and the Mind: An Introduction to Cognitive Science* (Harvard Univ Press, Cambridge, MA).
35. Hale JT (2014) *Automaton Theories of Human Sentence Comprehension* (CSLI Publications, Center for the Study of Language and Information, Stanford, CA).
36. Jackendoff R (1977) *X Syntax: A Study of Phrase Structure. Linguistic Inquiry Monograph 2* (MIT Press, Cambridge, MA).
37. Glasser MF, et al. (2016) A multi-modal parcellation of human cerebral cortex. *Nature* 536:171–178.
38. Vagharchakian L, Dehaene-Lambertz G, Pallier C, Dehaene S (2012) A temporal bottleneck in the language comprehension network. *J Neurosci* 32:9089–9102.
39. Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011) Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J Neurosci* 31:2906–2915.
40. Michel J-B, et al.; Google Books Team (2011) Quantitative analysis of culture using millions of digitized books. *Science* 331:176–182.
41. Devlin JT, Jamison HL, Matthews PM, Gonnerman LM (2004) Morphology and the internal structure of words. *Proc Natl Acad Sci USA* 101:14984–14988.
42. Lau EF, Phillips C, Poeppel D (2008) A cortical network for semantics: (De)constructing the N400. *Nat Rev Neurosci* 9:920–933.
43. Nakamura K, Dehaene S, Jobert A, Le Bihan D, Kouider S (2005) Subliminal convergence of Kanji and Kana words: Further evidence for functional parcellation of the posterior temporal cortex in visual word perception. *J Cogn Neurosci* 17:954–968.
44. Tuennerhoff J, Noppeney U (2016) When sentences live up to your expectations. *NeuroImage* 124:641–653.
45. Weber K, Lau EF, Stillerman B, Kuperberg GR (2016) The yin and the yang of prediction: An fMRI study of semantic predictive processing. *PLoS One* 11:e0148637.
46. McCauley SM, Christiansen MH (2011) Learning simple statistics for language comprehension and production: The CAPPUCCINO model. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pp 1619–1624.
47. McCarthy RA, Warrington EK (2013) *Cognitive Neuropsychology: A Clinical Introduction* (Academic, San Diego).
48. Mathy F, Feldman J (2012) What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition* 122:346–362.
49. Amalric M, et al. (2017) The language of geometry: Fast comprehension of geometrical primitives and rules in human adults and preschoolers. *PLOS Comput Biol* 13:e1005273.
50. Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C (2015) The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88:2–19.
51. Fitch WT (2014) Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Phys Life Rev* 11:329–364.
52. Abutalebi J, et al. (2007) The neural cost of the auditory perception of language switches: An event-related functional magnetic resonance imaging study in bilinguals. *J Neurosci* 27:13762–13769.
53. Piai V, et al. (2016) Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proc Natl Acad Sci USA* 113:11366–11371.
54. Hagen E, et al. (2016) Hybrid scheme for modeling local field potentials from point-neuron networks. *Cereb Cortex* 26:4461–4496, 10.1093/cercor/bhw237.
55. Miller KJ, Sorensen LB, Ojemann JG, den Nijs M (2009) Power-law scaling in the brain surface electric potential. *PLOS Comput Biol* 5:e1000609.
56. Eliasmith C, et al. (2012) A large-scale model of the functioning brain. *Science* 338:1202–1205.
57. Plate TA (1995) Holographic reduced representations. *IEEE Trans Neural Netw* 6:623–641.
58. Smolensky P (1990) Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artif Intell* 46:159–216.
59. Frankland SM, Greene JD (2015) An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proc Natl Acad Sci USA* 112:11732–11737.
60. Ray S, Maunsell JHR (2011) Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol* 9:e1000610.
61. Watanabe T, et al. (2013) A pairwise maximum entropy model accurately describes resting-state human brain networks. *Nat Commun* 4:1370.
62. Zaykin DV (2011) Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *J Evol Biol* 24:1836–1841.