Finding structure during incremental speech comprehension

Bingjiang Lyu[1*], William D. Marslen-Wilson[2], Yuxing Fang[2], Lorraine K. Tyler[2*]

[1]Changping Laboratory, Beijing, 102206, China

[2]Centre for Speech, Language and the Brain, Department of Psychology,

University of Cambridge, Cambridge, CB2 3EB, United Kingdom

* Correspondence: bingjiang.lyu@gmail.com (BL), lkt10@cam.ac.uk (LKT).

Abstract

A core aspect of human speech comprehension is the incremental combination of consecutive words into a structured and coherent interpretation of the speaker's intended meaning. This rapid process is subject to multi-dimensional probabilistic constraints, including both linguistic and non-linguistic knowledge in the specific context, and it is their interpretative coherence that drives successful comprehension. To unveil the neural substrates of this process, we extracted word-by-word measures of sentential structure from artificial neural networks, approximating a coherent outcome of the dynamic interplay between various types of constraints that is difficult to model with traditional methods. Using representational similarity analysis, we tested these structural measures and relevant lexical properties against the spatiotemporally resolved brain activity recorded by electro/magnetoencephalography when participants were listening to the same sentences. Our results reveal a detailed picture of the neurobiological processes involved in building structured interpretations through the integration across multifaceted constraints, including an extensive set of bilateral brain regions beyond the classical fronto-temporal language system, which sheds light on the distributed nature of language processing in the brain. This study also highlights the power of combining multiple methodologies to uncover the neural dynamics of complex cognitive processes.

Keywords: incremental speech comprehension, artificial neural networks, deep language models, representational similarity analysis, computational cognitive neuroscience, EEG/MEG

Introduction

Human speech comprehension involves a complex set of processes that transform an auditory input into the speaker's intended meaning, wherein each word is sequentially recognized and integrated with the preceding words to obtain a coherent interpretation[1-3]. Crucially, rather than simple linear concatenation, individual words are combined according to the nonlinear and often discontinuous structure embedded in an utterance as it is delivered over time[4]. For example, in the sentence *"The boy who chased the cat was…"*, it is the structurally close word *"boy"*, rather than the linearly close word *"cat"*, that is combined with *"was"*. However, the neural dynamics underpinning the incremental construction of a structured interpretation from a spoken sentence is still unclear.

Previous neuroimaging studies on the structure of language primarily focused on syntax[5], contrasting grammatical sentences against word lists or sentences with syntactic violations[6, 7], manipulating the syntactic complexity in sentences[8], or studying artificial grammatical rules elicited by structured nonsense strings[9]. Nevertheless, finding the structure in an unfolding sentence also depends on the constraints jointly placed by other linguistic properties and broad world knowledge[10, 11]. According to the widely accepted *constraint-based* approach to sentence processing[12, 13], human real-time interpretation of an utterance is subject to multiple types of probabilistic constraints (e.g., syntax, semantics, world knowledge) generated by individual words as they are sequentially heard in an spoken sentence, and where it is the *interpretative coherence* of these constraints that forms the basis for successful language comprehension[14]. Although lexical constraints of individual words can be estimated from large corpora data, it has historically been challenging to model the dynamic interplay between various types of linguistic and nonlinguistic information in a specific context, especially at the sentence level and beyond.

Contemporary deep language models (DLMs) have made great strides in a wide array of natural language processing tasks, including text generation, parsing and translation[15-18]. While current DLMs are still imperfect in terms of human-level language understanding related to reasoning and complex physical and social situations[19], they are arguably valuable models of general linguistic capacities, due to their ability to identify and leverage relevant statistical regularities of linguistic and non-linguistic knowledge present in massive training data[20-22]. Human language comprehension seems to require an apparently analogous contextualized integration of multifaceted constraints[11, 23, 24]. In this regard, DLMs excel in flexible combination of different types of features embedded in their rich internal representations[25]. Their deep contextualized representations capture the distributed regularities that jointly determine the coherent interpretation of a given sentence, providing context-dependent composition and quantitative measures of the underlying sentential structure. These properties relate back to Elman's recurrent neural network[26, 27] which automatically picks up and encodes lexical syntactic/semantic information in the hidden states.

Recent studies have revealed an overall congruence between language representations in DLMs and those observed in the human brain while processing the same spoken or written

input[28-34], suggesting the potential value of DLMs as a computational tool to investigate the neural basis of language comprehension. To move beyond comparing the similarities between entire model hidden states and brain activity, probing techniques that can extract specific contents from DLMs[35, 36] make it possible to study the neural dynamics relevant to processing such information. The important advance here is that we can leverage the deep learning strengths of DLMs to create rigorously quantified models of the broader and multifaceted constraint environment in which a structured interpretation is constructed. Such models can be compared, dynamically, with more restricted and interpretable factors that capture the specific linguistic combinatorial constraints necessary for successful language comprehension.

Here, we take this approach further by designing sentences with contrasting linguistic structures and using a structural probe technique[37] to extract word-by-word contextualized representations of sentential structures from a widely-used DLM, namely, BERT[16]. This provides the neurocomputational model specificity required to elucidate the neural dynamics underlying the online construction of a structured interpretation from an unfolding spoken sentence. After a detailed evaluation of BERT structural measures according to the hypothesized *constraint-based* approach and human behavioral results, we used spatiotemporal searchlight representational similarity analysis (ssRSA)[38] to test these quantitative structural measures and relevant lexical properties against source-localized EMEG data recorded while participants were listening to the same sentences. These tests reveal how the structured interpretation of a spoken sentence is incrementally built under multifaceted probabilistic constraints in the brain.

Results

We constructed 60 sets of sentences with varying sentential structures (see Methods) and presented them to human listeners. We also input them word-by-word to BERT to extract incremental structural representations. These natural spoken sentences were constructed to balance off specifically linguistic constraints on interpretation against varying non-linguistic constraints as the sentence is incrementally interpreted, providing a realistic simulation of the environment of daily language use. In each stimulus set, there are two target sentences differing only in the transitivity of the first verb (Verb1) encountered, i.e., how likely it is that Verb1 takes a direct object [see (1) and (2) below and Fig. 1A]:

(1) *The dog found in the park was covered in mud.*
(2) *The dog walked in the park was covered in mud.*

In the first sentence, Verb1 (i.e., *"found"*) has high transitivity (HiTrans) and strongly prefers a direct object (e.g., ball), while in the second sentence, Verb1 (i.e., *"walked"*) has relatively low transitivity (LoTrans) and is often used without a following direct object. Critically, (a) the structural interpretation of these sentences is ambiguous at the point Verb1 is encountered and (b) the preferred human resolution of this ambiguity depends on the real-time integration of linguistic and non-linguistic probabilistic constraints as more of the sentence is heard. In the

example above, the sequence *"The dog found...*" could initially have either an Active interpretation – where the dog has found something, or a Passive interpretation – where the dog is found by someone (Fig. 1B). Because *"find"* is primarily a transitive verb, the human listener is likely to be biased towards an initial Active interpretation. Similarly, the sequence *"The dog walked…"*, where *walk* is primarily used as an intransitive verb (without a direct object), could also bias the listener to an Active interpretation, where the dog is doing the walking, rather than the less frequent Passive interpretation where someone is taking the dog for a walk (i.e., walking the dog).

This initial structural interpretation up to Verb1 does not, however, just depend on linguistic knowledge such as Verb1 transitivity. It also depends on how likely the subject is (or is not) to adopt the Active (agent) role to perform the specified action[39, 40], that is, "thematic role" properties of the subject noun. This likelihood, of the event structure implied by the different structural combinations of the subject noun and Verb1, will depend on wide ranging knowledge of the world, linked to the specific words being heard. So, regardless of Verb1 transitivity, the Active interpretation should be more strongly favored in *"The king found/walked..."* given the higher agenthood of the *"king"* and thus the greater implausibility of a Passive interpretation involving a *"king"* relative to a *"dog"*. Hence, the word-by-word interpretation of the sentential structure – and of the real-world event structure evoked by this interpretation – is determined by the constraints jointly placed by the subject noun and Verb1, which is manifested by the interpretative coherence between linguistic knowledge and world knowledge.

As the sentence evolves, and the prepositional phrase *"in the park"* that follows Verb1 is incrementally processed, there is further modulation of the preferred interpretation, again reflecting both Verb1 transitivity and the plausibility of the event being constructed. Specifically, the Passive interpretation will become more preferred in a HiTrans sentence, given the absence of an expected direct object for the highly transitive Verb1, so Verb1 tends to be interpreted as a passive verb [i.e., the head of a reduced relative clause in *"The dog (that was) found in the park…"*]. Conversely, in a LoTrans sentence, the Active interpretation of Verb1 is strengthened by the incoming prepositional phrase, which is in accord with the verb's intransitive use and the event conjured up by the sequence of words heard so far (e.g., *"The dog walked in the park…"*). Hence, these two sentence types are likely to differ in the structural interpretation preferred by the end of the prepositional phrase. However, with the appearance of the actual main verb (e.g., *"was covered"* in the example sentences), the Active interpretation of Verb1 as the main verb will be completely rejected, which resolves the potential ambiguity and confirms the Passive interpretation in both HiTrans and LoTrans sentences.

In brief, understanding these complex sentences require listeners to integrate discontinuous words to solve a long-distance dependency between the subject noun and the actual main verb separated by an intervening clause. This engages the neurobiological processes of integration across multiple levels of the sentence processing system and different lexical constraints, for example, the incremental building, maintenance and update of sentential structure over time might primarily involve activity in the fronto-temporal regions[41], while estimating the

plausibility of the event interpreted from the sentence with prior knowledge of the world may elicit neural responses in the default mode network[42].

Human incremental structural interpretations

As the first step, and to quantify how the stimulus sentences exemplified a constraint-based account of incremental structural interpretation, we conducted two pre-tests where participants listened to sentence fragments, starting from sentence onset and continuing either until the end of Verb1 or to the end of the prepositional phrase (Fig. 1A), and then produced a continuation to complete the sentence (see Methods). Based on the continuations provided by the listeners at these two gating points, we can infer their online structural interpretations.

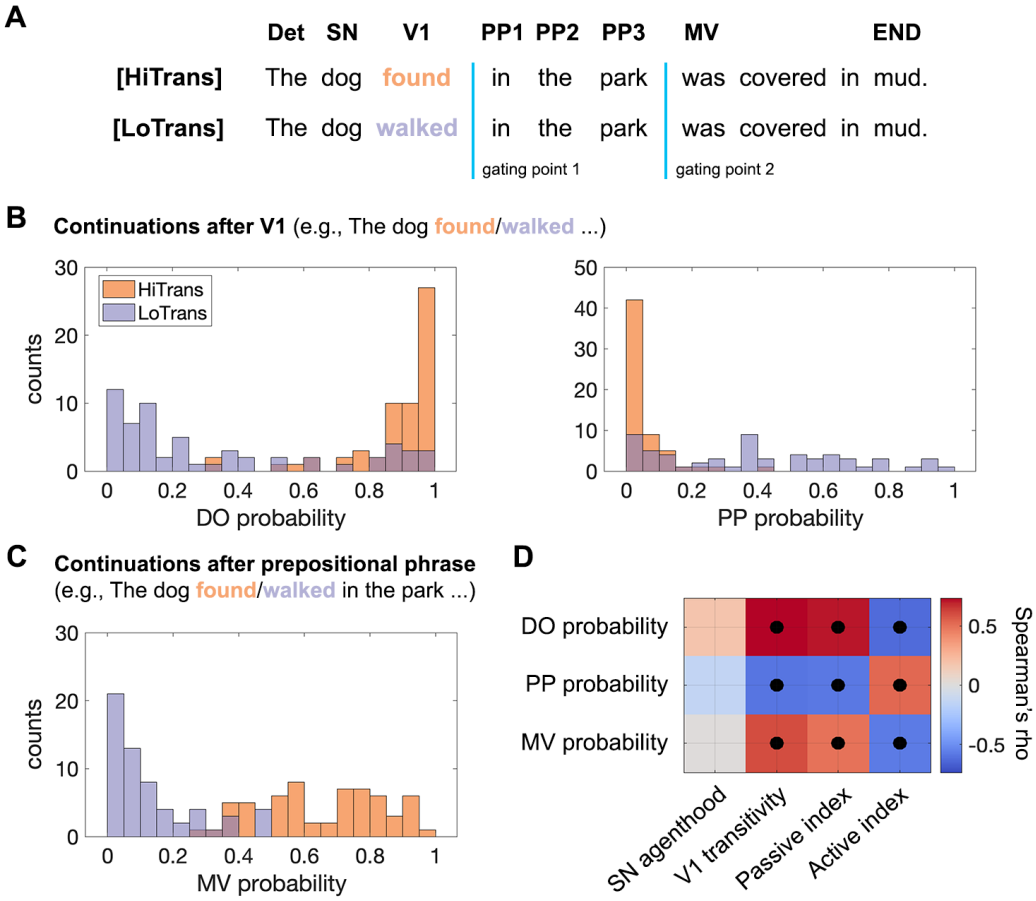

Fig. 1. Human incremental structural interpretations derived from continuation pre-tests. **(A)** An example set of target sentences differing only in the transitivity of Verb1, HiTrans: high transitivity, LoTrans: low transitivity. Det: determiner, SN: subject noun, V1: Verb1, PP1-PP3: prepositional phrase, MV: main verb, END: the last word in the sentence. **(B)** Probability of a direct object (left) and a prepositional phrase (right) in the continuations after Verb1. **(C)** Probability of a main verb in the continuations after Verb1, which indicates an Active interpretation. **(D)** Correlations between multifaceted lexical constraints and probabilistic interpretations in the two pre-tests. (Spearman rank correlation, black dots indicate significance determined by 10,000 permutations, $P_{FDR} < 0.05$ corrected).

In the continuations after Verb1, a direct object was more likely to be found in HiTrans sentences, indicating a transitive use of Verb1, while an opposite pattern was found for a PP

continuation, indicating an intransitive use of Verb1 (Fig. 1B). As expected, the probability of a main verb (MV) in the continuations after the prepositional phrase was lower in LoTrans sentences (Fig. 1C), suggesting that listeners preferred the Active interpretation and tended to interpret Verb1 as the main verb by the end of the prepositional phrase in LoTrans sentences, and vice versa in HiTrans sentences.

Crucially, neither of the two pre-tests resulted in a complete separation between HiTrans and LoTrans sentences; instead, they were characterized by two different but overlapping probabilistic distributions. This suggests that Passive and Active interpretations varied in plausibility in each sentence type before the actual main verb was presented, reflecting the probabilistic constraints jointly placed by the combination of the specific subject noun, Verb1, and the prepositional phrase in each sentence.

To relate these human interpretative preferences to the broader landscape of distributional language data, we developed corpus-based measures of the thematic role preference of the subject noun (i.e., how likely it is interpreted as an agent that conducts an action) and the transitivity of Verb1 in each sentence, from which we derived a Passive index and an Active index. These indices separately capture the interpretative coherence between these two types of lexical properties towards Passive and Active interpretations (see Methods). Both high subject noun agenthood and low Verb1 transitivity coherently preferred an Active interpretation as the prepositional phrase was heard (i.e., a high Active index), and vice versa for the Passive interpretation (i.e., a high Passive index). In accord with the *constraint-based* hypothesis, we found that human interpretative preference for the two types of sentences was significantly correlated with the lexical constraints generated by the subject noun and Verb1 (Fig. 1D).

Incremental structural representations extracted from BERT

Next, we extracted structural representations at various positions in the same sentences from BERT and evaluated them according to the *constraint-based* hypothesis and human behavioral results. This motivates the use of BERT structural measures to reveal how the structured interpretation of a spoken sentence is incrementally built in the brain.

Typically, the structure of a sentence can be represented by a dependency parse tree[43] (Fig. 2A) where words are situated at different depths given their structural dependency. Each edge links two structurally proximate words as being the head and the dependent separately (e.g., a verb and its direct object). However, such a parse tree is context-free, that is, it only captures the syntactic relation between each pair of words and abstracts away from the specific lexical (and higher order) contents of the sentence that constrain its actual online structural interpretation. This context-free parse depth is always the same for words at the same position in sentences with the same structural interpretation (e.g., *"found"* and *"walked"* in either of the two parse trees in Fig. 2A).

To obtain structural measures that also encode the specific lexical and higher order contents in a sentence, we adopted a structural probing technique[37] to reconstruct a sentence's structure by

estimating each word's parse depth based on their contextualized representations generated by BERT (see Methods). Note that BERT is a multi-layer DLM (24 layers in the version used in this study) which may distribute different aspects of its computational solutions over multiple layers. Accordingly, we trained a structural probing model for each layer, and selected the one with the most accurate structural representations while also including its neighboring layers to cover relevant upstream and downstream information. Following this strategy, we used the BERT structural measures obtained from layers 12-16 with the best performance achieved in layer 14 (see Fig. S1 and Methods).

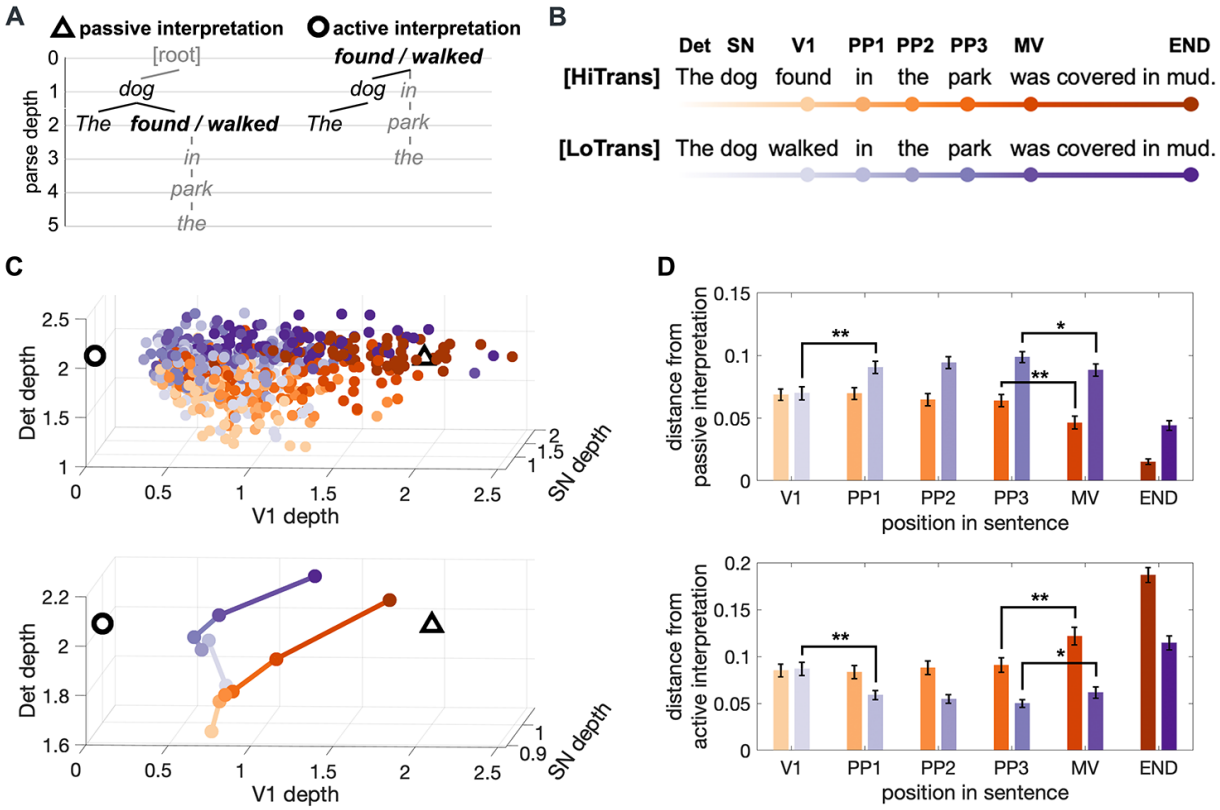

Fig. 2. Incremental interpretation of sentential structure by BERT. (A) Context-free dependency parse trees of two plausible structural interpretations. Left: Passive interpretation where V1 is the head of a reduced relative clause. Right: Active interpretation where V1 is the main verb. **(B)** Incremental input to BERT, with the lightness of dots encoding different positions in the target sentences. Det: determiner, SN: subject noun, V1: Verb1, PP1-PP3: prepositional phrase, MV: main verb, END: the last word in the sentence. **(C)** Incremental interpretations of the dependency between SN and V1 in the model space consisting of the parse depth of Det, SN and V1. Upper: Each colored circle represents the parse depth vector up to V1 derived at a certain position in the sentence [with the same color scheme as in (A)]. The hollow triangle and circle represent the context-free dependency parse vectors for Passive and Active interpretations in (B). Lower: incremental interpretations of the two types of target sentences represented by the trajectories of median parse depth. **(D)** Distance from Passive and Active landmarks in the model space as the sentence unfolds [between each colored circle and the two landmarks in the upper panel of (C)] (two-tailed two-sample t-test, *: $P < 0.05$, **: $P < 0.001$, error bars represent SEM).

We input each sentence word-by-word to the trained BERT structural probing models, focusing on the incremental structural representation being built as it progressed from Verb1 to the main verb (see the sequence in Fig. 2B). Note that we defined the first word after the prepositional phrase as the main verb since its appearance is sufficient to resolve the intended structure where Verb1 is a passive verb. We found that, for each type of sentences, the BERT parse depth of words at the same position formed a distribution ranging around the corresponding context-free parse depths in either the Passive or the Active interpretation (see Fig. S2), suggesting a word-specific rather than position-specific structural representation.

Next, we visualized BERT's word-by-word structural measures, focusing on the dependency between the subject noun and Verb1 that is core to the current interpretation of the sentence – whether the subject noun is the agent or the patient of Verb1. To this end, we built a 3-dimensional vector including BERT parse depths of the first three words up to Verb1 for each sentence (e.g., *"The dog found..."*). This 3D vector was kept updated every time the input increased by one word in length, capturing the dynamic interpretation of the structural dependency between the subject noun and Verb1 given the contents of the subsequent words in a specific sentence. Similar to the probabilistic interpretation found within each type of sentences in human listeners, trajectories of individual HiTrans and LoTrans sentences are considerably distributed and intertwined (Fig. 2C, upper), suggesting that BERT structural interpretations are sensitive to the idiosyncratic contents in each sentence.

To make sense of these trajectories, we also vectorized the context-free parse depth of the first three words indicating Passive and Active interpretations separately and located them in the 3D vector space as landmarks (hollow triangle and circle in Fig. 2C), so that the plausibility of either interpretation can be estimated by a sentence's distance from the corresponding landmark. As shown by the trajectories of the median BERT parse depth of the two sentence types (Fig. 2C, lower), in general, HiTrans sentences continuously moved towards the Passive interpretation landmark after Verb1, with a significant change of distances detected at the main verb (Fig. 2D, orange bars). LoTrans sentences started by approaching the Active interpretation landmark but were reorientated to the Passive counterpart with the appearance of the actual main verb, with significant changes of distances detected at both Verb1 and main verb (Fig. 2D, purple bars). These results resemble the pattern of human interpretative preference observed in the continuation pre-tests, where the Passive and Active interpretations were separately preferred in HiTrans and LoTrans sentences by the end of the prepositional phrase in a probabilistic manner (Fig. 1), before the Passive interpretation was established with the appearance of the actual main verb.

BERT structural measures are correlated with constraints driving human interpretation

Moreover, similar to human listeners, we found that BERT's preference for structural interpretation was also correlated with the constraints placed by the subject noun and Verb1 (see Methods). We first focused on BERT's interpretative mismatch quantified as the distance between an unfolding sentence and each of the two landmarks in the model space, which was dynamically updated as the sentence unfolded (Fig. 2C). Consistently, from the incoming prepositional phrase to the main verb, sentences that are closer to the Passive landmark in the

vector space have higher Verb1 transitivity, a higher Passive index but a lower Active index, sentences closer to the Active interpretation landmark exhibited higher Active index and lower Passive index (Figs. 3A and 3B). Moreover, at the beginning of the prepositional phrase, the change of distance towards either interpretation landmark between two consecutive words is also correlated with these constraints (Figs. 3C and 3D), suggesting an immediate update in the structural interpretation in combination with the accumulated constraints from the preceding subject noun and Verb1.

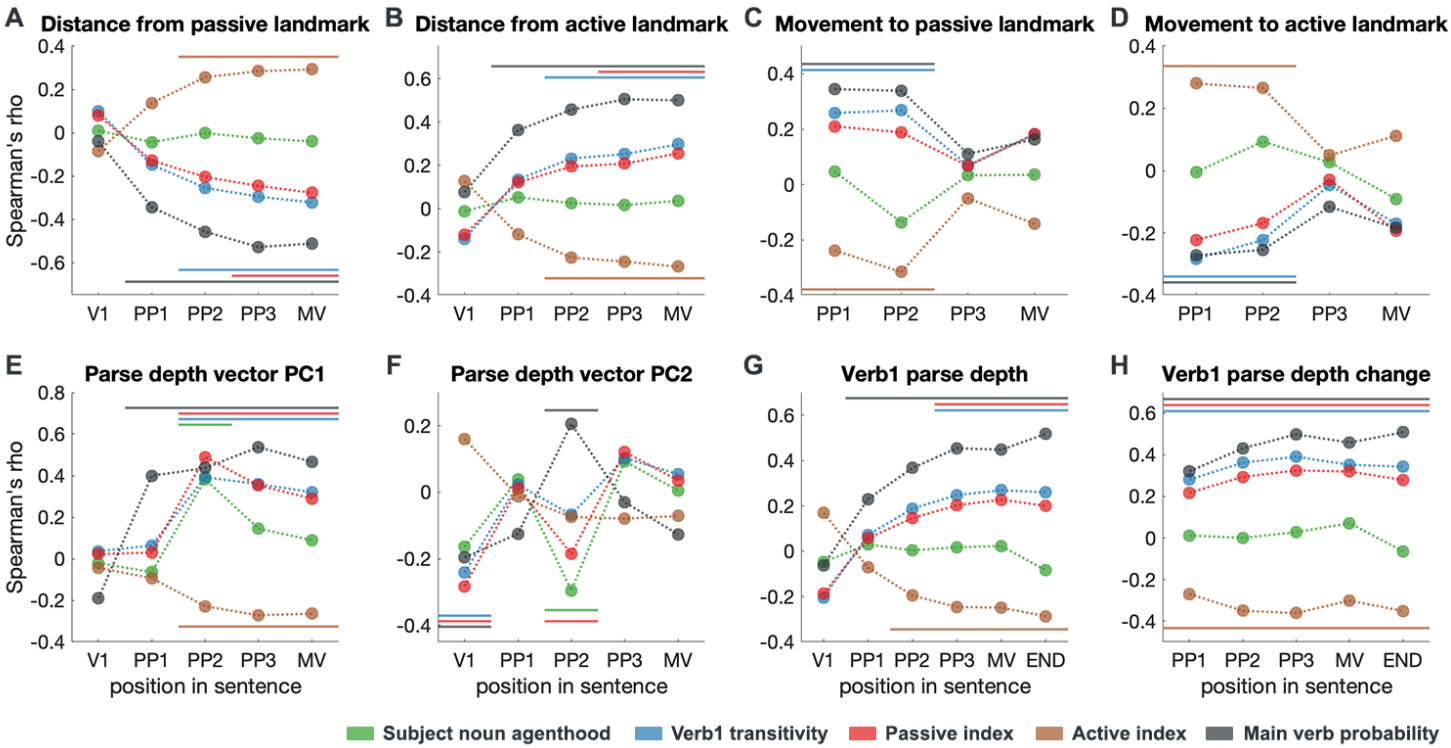

Fig. 3. Correlation between incremental BERT structural measures and explanatory variables. BERT structural measures include **(A, B)** BERT interpretative mismatch represented by each sentence's distance from the two landmarks in model space (Fig. 1C); **(C, D)** Dynamic updates of BERT interpretative mismatch represented by each sentence's movement to the two landmarks; **(E, F)** Overall structural representations captured by the first two principal components (i.e., PC1 and PC2) of BERT parse depth vectors; **(G, H)** BERT Verb1 (V1) parse depth and its dynamic updates. Explanatory variables include lexical constraints derived from massive corpora and the main verb probability derived from human continuation pre-tests (Spearman correlation, permutation test, $P_{FDR} < 0.05$, multiple comparisons corrected for all BERT layers, results shown here are based on layer 14, see Figs. S3-S5 for the results of all layers); PP1-PP3: prepositional phrase, MV: main verb, END: the last word in the sentence.

Similarly, we found that both the incremental BERT parse depth vectors as a whole (which are captured by their principal components) and the BERT parse depth of Verb1 (which is the most indicative marker of the interpretation preferred) are correlated with the constraints placed by the subject noun and Verb1 (Figs. 3E to 3H). Moreover, the significant effects consistently found as the sentence unfolds suggest that properties of preceding words are used to constrain the interpretation of the upcoming input, which is key to resolving discontinuous structural

dependencies. In addition, we found that BERT structural interpretations were also correlation with the main verb probability in the continuation pre-test which directly reflects human interpretation preferences (black bars in Fig. 3).

Overall, these results illustrated, at which position in a sentence, relevant lexical constraints started being encoded by BERT, which also validated the contextualized BERT structural measures in terms of the *constraint-based* hypothesis and human behavioral results, and motivated the use of them to probe the neural processes involved during the incremental interpretation of sentence structure.

Neural dynamics of incremental structural interpretation

To study how the structured interpretation of a spoken sentence is built word-by-word in the brain, we used ssRSA to test the incremental BERT structural measures in source-localized EMEG collected when the same sentences were delivered to human listeners. This combination of methods gains improved neurocomputational specificity by probing the spatiotemporally resolved neural activity with detailed structural representations rather than the entire hidden states. We compared the representational geometry of BERT structural measures with that of neural responses inside a spatiotemporal searchlight moving across the brain, significant similarity fits showed when and where the incremental structural interpretations emerge and update in the brain. Given the probabilistic interpretations in BERT and human listeners reported above, we combined HiTrans and LoTrans sentences as one group to increase the range of pair-wise dissimilarity to be modelled in RSA.

We began with the BERT parse depth vector containing the parse depth of each word in an incremental input, providing a dynamic structural representation updated as the sentence unfolded. Then, we tested the interpretative mismatch between the incremental BERT parse depth vector and the corresponding context-free parse depth vector for the Passive or the Active interpretation. The degree of this mismatch is proportional to the evidence for or against the two interpretations, i.e., the smaller the distance, the more positively loaded this interpretation. Besides these two measures based on the entire incremental input, we also focused on Verb1 since the potential structural ambiguity lies in whether Verb1 is interpreted as a passive verb or the main verb. Given the context-free parse depth of Verb1 that is 2 in the passive interpretation and 0 in the Active interpretation (Fig. 2B), with each incoming later word, an increased BERT Verb1 parse depth towards 2 or a decreased value towards 0 reflects separately the preference biased to a Passive or an Active interpretation (Fig. S6, see Table S1 for a summary of all BERT measures tested).

For the listener's neural activity, we focused on three critical epochs in each sentence: (a) Verb1 – when its structural dependency with the preceding subject noun was initially established despite potential ambiguity, (b) the preposition – when the initial structural interpretation started being updated, to be either strengthened or weakened by the incoming preposition phrase, and (c) main verb – when the intended Passive interpretation was finally confirmed. We aligned the continuous EMEG data to the onset of Verb1, the preposition, and the main verb respectively and obtained three 600-ms epochs.

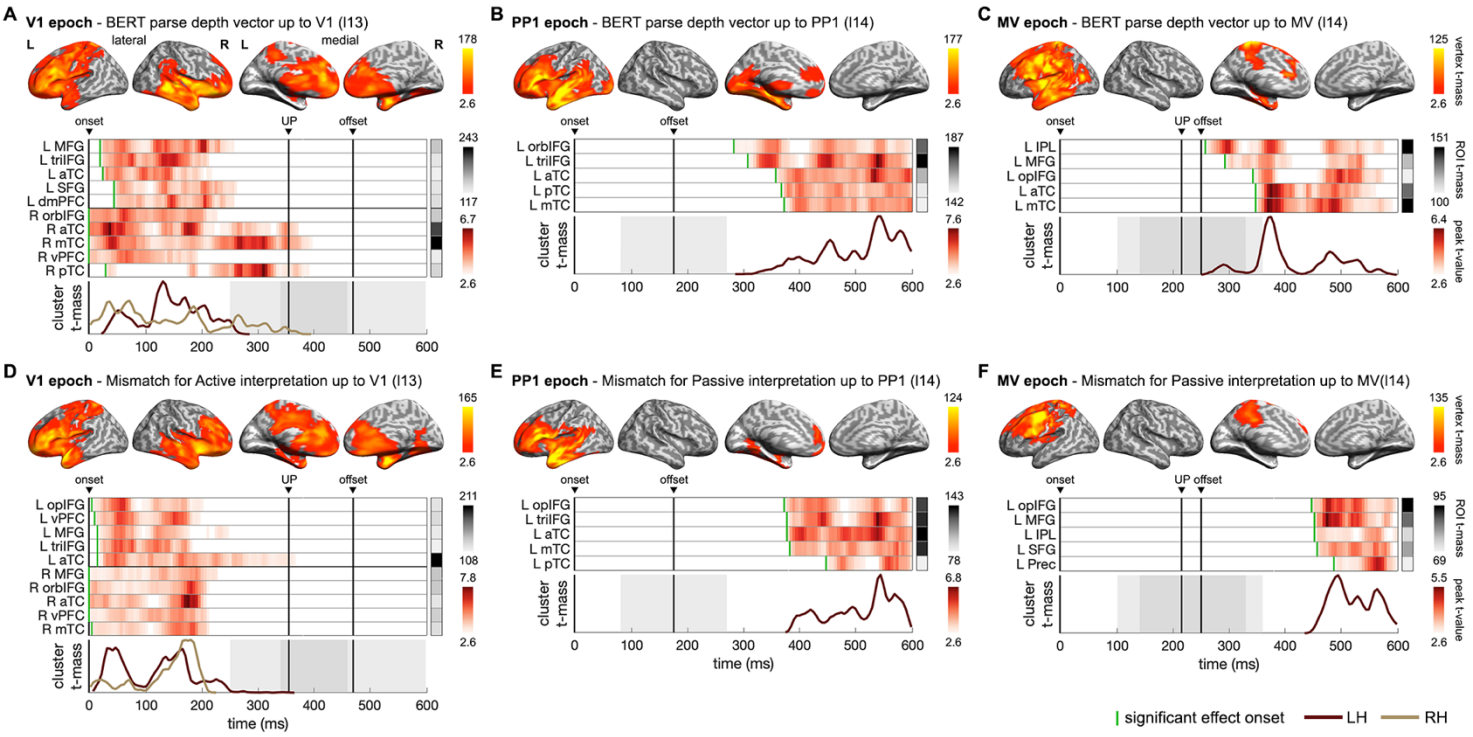

Fig. 4. Neural dynamics underpinning the emerging structure and interpretation of an unfolding sentence. (A-C) ssRSA results of BERT parse depth vector up to Verb1 (V1), the preposition (PP1) and the main verb (MV) in epochs separately time-locked to their onsets. **(D-F)** ssRSA results of the mismatch for the preferred structural interpretation (the specific BERT layer from which BERT structural measures were derived was denoted in parentheses). From top to bottom in each panel: vertex t-mass (each vertex's summed t-value during its significant period); heatmap of time-series of ROI peak t-value (the highest t-value in an ROI at each time-point) with a green bar indicating effect onset and ROI t-mass (each ROI's summed mean t-value during its significant period); cluster t-mass time-series (summed t-value of all the significant vertices of a cluster at each time-point). [cluster-based permutation test, vertex-wise $P < 0.01$, cluster-wise $P < 0.05$ in (A-E); marginally significance in (F) with cluster-wise $P = 0.06$]. Solid vertical lines indicate the timings of onset, average uniqueness point (UP), and average offset of the word time-locked in the epoch with grey shades indicating the range of one SD. LH/RH: left/right hemisphere. See Table S2 for full anatomical labels. See Fig. S8 for the significant results of other BERT layers in the MV epoch.

We found that the incremental BERT parse depth vectors exhibited significant fits to brain activity consistently in all three epochs, as the corresponding word was being heard at that time (Figs. 4A to 4C). In Verb1 epoch, effects in bilateral frontal and anterior-to-middle temporal regions started immediately from Verb1 onset and continued until the uniqueness point – the point at which the word has been uniquely identified – while the BERT parse depth of Verb1 *per se* showed similar but with greater duration which peaked exactly at Verb1 uniqueness point (Fig. S7). As the sentence unfolded, effects were found in the left fronto-temporal regions in the two later epochs, starting after the recognition of the preposition or the main verb separately.

Turning to the interpretative mismatch for the two possible interpretations, we only observed significant effects of the mismatch for Active interpretation in Verb1 epoch (Fig. 4D). However, it was the mismatch for Passive interpretation that fitted brain activity in the preposition and main verb epochs (Figs. 4E and 4F, marginal significance in main verb epoch with cluster-wise P = 0.06). These results suggest that listeners, in general, tended to have an initial preference for an Active interpretation but might start favoring a Passive interpretation when the prepositional phrase began to be heard. This finding is consistent with the tendency to process the first noun encountered in a sentence as the agent[10, 44].

Effects of the BERT parse depth vectors and those of the interpretative mismatch for the preferred structural interpretation have substantial overlaps in terms of their spatiotemporal patterns in the brain, characterized primarily by a transition from bilateral to left-lateralized fronto-temporal regions as the sentence unfolds. Across the three epochs, the most sustained effects were observed in the left inferior frontal gyrus (IFG) and the anterior temporal lobe (ATL). Notably, with the identification of the actual main verb, effects of the eventually resolved structure also involved regions in the left prefrontal and inferior parietal regions (Fig. 4C) which belong to the multiple-demand network[45].

Structural ambiguity resolution probed using BERT Verb1 parse depth

As mentioned above, the potential ambiguity between a Passive and an Active interpretation centers around whether Verb1 is considered as a passive verb or the main verb, which is resolved upon the appearance of the actual main verb. We probed how this is implemented in the brain using the dynamic BERT parse depth of Verb1. Specifically, the cognitive demands required by this resolution process can be characterized by the change between the updated BERT parse depth of Verb1 when the actual main verb is presented and its initial value when Verb1 is first encountered (see Fig. S6 for the dynamic change of BERT V1 parse depth).

We first tested the change of Verb1 parse depth in the main verb epoch. Significant fits to brain activity emerged in the left posterior temporal and inferior parietal regions upon the main verb uniqueness point, and then extended to more anterior temporal regions (Fig. 5A). After the main verb offset, the declining effects of the Verb1 parse depth change in the left anterior temporal region seamlessly overlapped with the arising effects of the updated Verb1 parse depth (Figs. 5B and 5C). These results indicate that the recognition of the actual main verb immediately triggered an update of the previous interpretation of Verb1, with the resolved interpretation emerged in the left temporal lobe and was later delivered to the right posterior temporal and parietal areas. It is also worth noting that the left hippocampus was activated for both measures of Verb1 parse depth after the actual main verb is recognized, suggesting that the episodic memory of experienced events might contribute to the updating of structural interpretations (46, 51, 52).

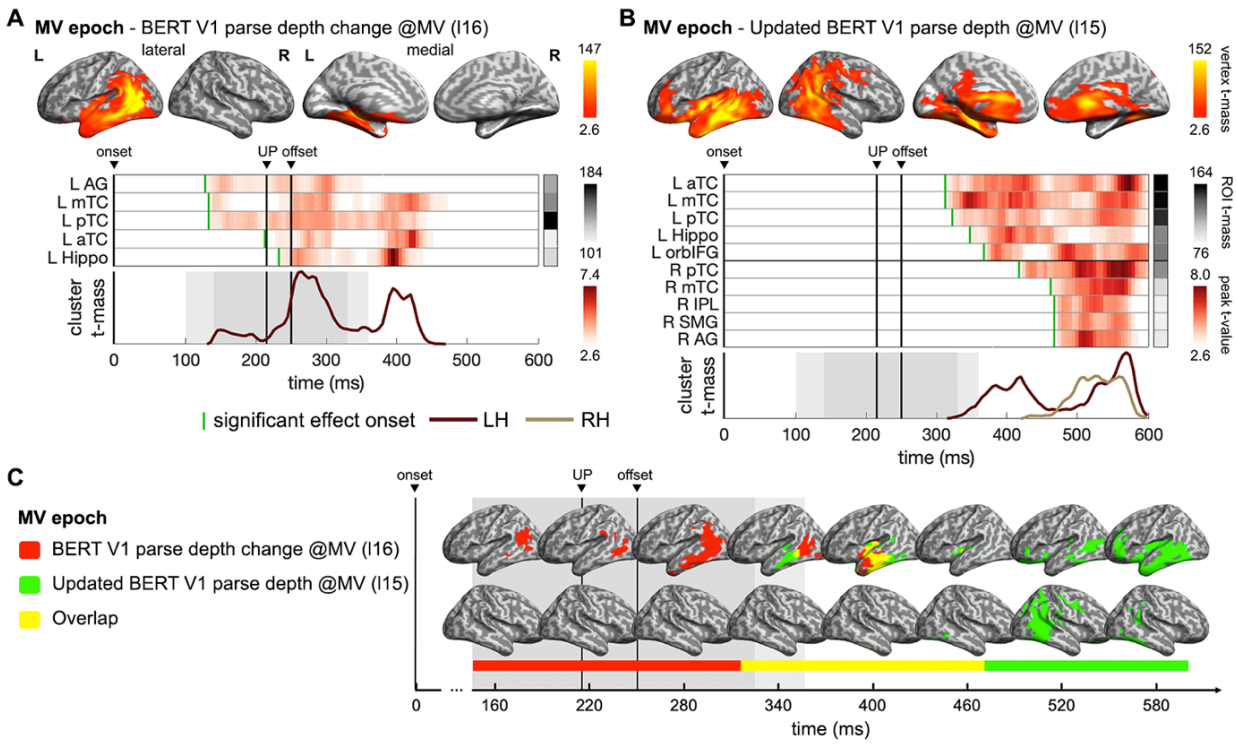

Fig. 5. Neural dynamics updating the incremental structural interpretation. (A) ssRSA results of BERT Verb1 (V1) parse depth change at the main verb (MV) relative to the parse depth V1 when it is first encountered. **(B)** ssRSA results of the updated BERT V1 parse depth when the input sentence reaches MV. **(C)** Spatiotemporal overlap between the effects in (A) and (B). (cluster-based permutation test, vertex-wise $P < 0.01$, cluster-wise $P < 0.05$).

Emergent structural interpretations driven by multifaceted constraints in the brain

Next, we further asked how the multifaceted constraints, which are also incorporated into BERT structural measures, drive the interpretation made by human listeners. When and where in the brain do these constraints emerge? How are their neural effects related to those of the final resolved sentential structure? To address these questions, we first tested the subject noun thematic role properties. Significant effects of agenthood and patienthood were found in the preposition epoch (Fig. 6A) and in the main verb epoch (Fig. 6B) separately. Notably, effects of the subject noun itself preceded those of incremental BERT parse depth vectors modelling the sentence fragments in the same epoch (compare Fig. 6A with Fig. 4B, and Fig. 6B with Fig. 4C). This indicates that subject noun thematic role might be evaluated before building the overall structural interpretation of the utterance delivered so far. Specifically, the initial preference for an Active interpretation during Verb1, while present as the prepositional phrase started (Fig. 6A), was superseded by the preference for a Passive interpretation as the rest of the phrase (Fig. 4E) and the main verb (Fig. 6B) were heard.

Despite being jointly constrained by subject noun thematic role preference and Verb1 transitivity in a probabilistic manner, the structural interpretation temporarily held just before the recognition of the actual main verb could differ across sentences (e.g., Passive interpretation in "*The dog found in the park*..." and Active interpretation in "*The dog walked in the park*..."). Therefore, in contrast to the Passive or Active index specialized for one

particular structural interpretation, we constructed a non-directional index that merely quantifies the degree of interpretative coherence for one interpretation, whether Passive or Active (see Methods). Thus, a higher value only indicates greater interpretative coherence between the subject noun and Verb1 regardless of which interpretation is preferred.

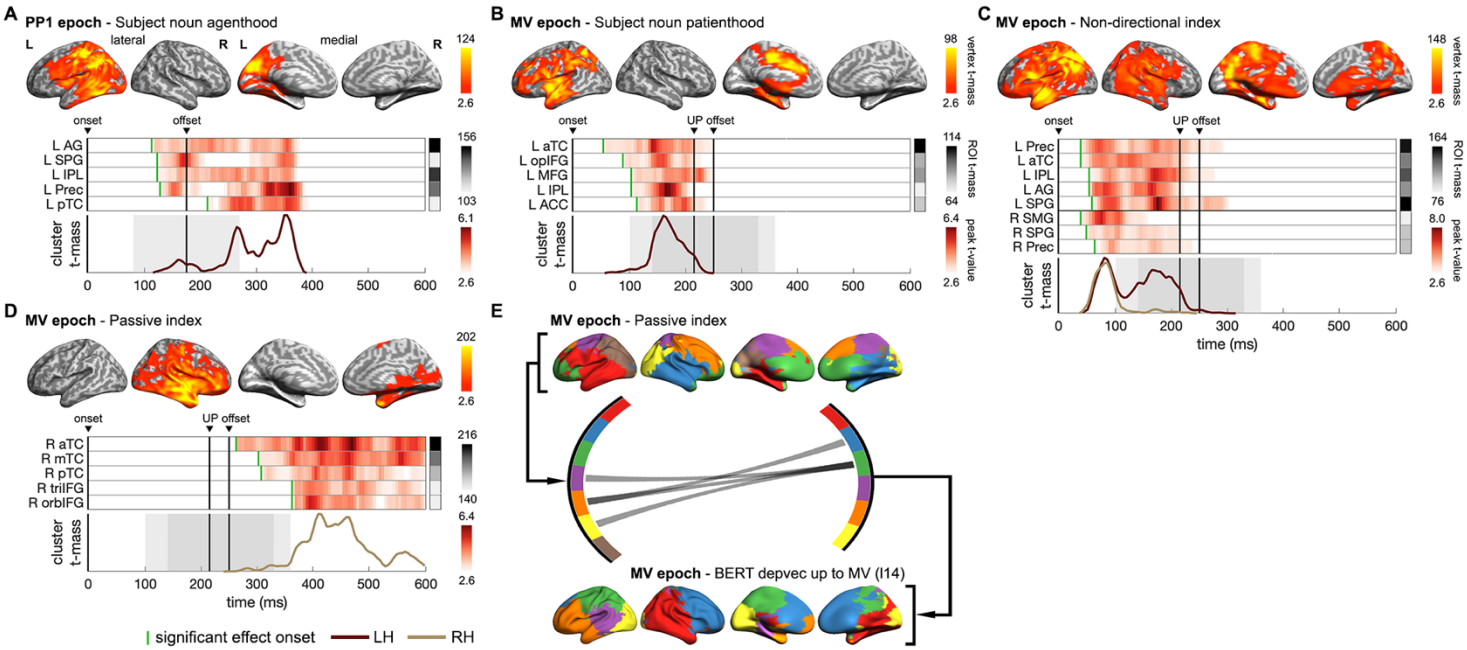

Fig. 6. Neural dynamics of multifaceted probabilistic constraints underpinning incremental structural interpretations. (A, B) ssRSA results of SN agenthood and SN patienthood (i.e., plausibility of SN being the agent or the patient of V1) in PP1 and MV epochs separately. **(C)** ssRSA results of non-directional index (i.e., interpretative coherence between SN and V1 regardless of the structure preferred) in MV epoch. **(D)** ssRSA results of Passive index (i.e., interpretative coherence for the Passive interpretation) in MV epoch. **(E)** Influence of the Passive interpretative coherence on the emerging sentential structure in MV epoch revealed by the Granger causal analysis (GCA) based on the non-negative matrix factorization (NMF) components of whole-brain ssRSA results (see Fig. S9 for more details) [(A-D) cluster-based permutation test, vertex-wise $P < 0.01$, cluster-wise $P < 0.05$; (E) permutation test $P_{FDR} < 0.05$].

Effects of this non-directional measure of interpretative coherence appeared very soon after the main verb onset in both hemispheres and lasted till its offset (Fig. 6C), suggesting an immediate evaluation of the previously integrated constraints from the subject noun and Verb1 when a listener realized that the sentence had not finished yet. Moreover, these effects roughly co-occurred with the effects of subject noun patienthood (compare Figs. 6B and 6C), indicating that a patient role for the subject noun was considered as the main verb was being recognized. Intriguingly, the most sustained regions associated with this non-directional index, including the left ATL, angular gyrus (AG) and precuneus, are also the classical areas of the default mode network (DMN). This finding is consistent with recent claims that the DMN integrates external input with internal prior knowledge to make sense of an input stimulus such as speech[42]. In particular, precuneus and AG have been found to be involved in building thematic relationships and event structures from episodic memory[46, 47].

Following the declining effects of the non-directional index upon the recognition of the main verb, we found significant effects of the Passive index in right anterior fronto-temporal regions (Fig. 6D), suggesting that the intended Passive interpretation was eventually established in all sentences. Previous studies have revealed that the relatively narrow sentence-specific information and the broad world knowledge are processed in the left and right hemispheres separately[48-50]. Relevant to this, in the main verb epoch, we found effects of the BERT parse depth vector and those of the Passive index in the left and right hemispheres respectively, arising almost at the same time as the main verb was recognized (compare Fig. 4C with Fig. 6D). Therefore, a critical question is whether and how the online structural interpretation of a specific sentence is facilitated by the interpretative coherence conjured up from lexical constraints that also depend on broad world knowledge (e.g., thematic role).

To address this question, we adopted non-negative matrix factorization to decompose the whole-brain RSA fits of the Passive index and the BERT parse depth vector found in the main verb epoch into two sets of components given their temporal synchronizations (see Methods). We then conducted multivariate Granger causality analyses (MGCA) to infer directed connections among them. We found only GC connections from the components of Passive index to those of BERT parse depth vector (Fig. 6E). Specifically, we identified information flows from the right hemisphere components of the Passive index to the left hemisphere components of BERT parse depth vector, suggesting that a specific sentence's structure represented in the left hemisphere might be influenced by the coarse estimate of the event plausibility concurrently determined by broad world knowledge in the right hemisphere[48] (see Fig. S9 for more details).

Discussion

In this study, we investigated the neural dynamics involved in constructing structured interpretations of spoken sentences on a word-by-word basis. We combined spatiotemporally resolved brain activity of human listeners, quantitative structural representations derived from a DLM (i.e., BERT), and measures of lexical constraints estimated from corpora data. Our study revealed the emergence and update of a structured interpretation, jointly constrained by various lexical properties related to both linguistic and non-linguistic knowledge, in an extensive set of brain regions beyond the core fronto-temporal language network. These findings provide empirical evidence for the constraint-based approach to sentence processing and deepen the understanding of specific spatiotemporal patterning and neuro-computational properties underpinning incremental speech comprehension.

Using artificial neural networks (ANNs) to study the neural substrates of human cognition complements the long-time pursuit of generative rules and interpretable models[51]. ANNs have informed our understanding of various cognitive processes in the brain by providing quantifiable predictions that aim to connect behaviors and relevant neural activity[52-59]. This is crucial for quantifying the outcome of complex, interrelated constraints that arise in specific

contexts, such as spoken sentences, and constructing the representational geometry to be probed in the brain. Where DLMs are concerned, recent studies have systematically compared the internal representations of DLMs to those observed in the human brain during language processing, which highlights the importance of predictive coding and contextual information[28-34]. Furthermore, these studies have motivated the use of DLMs as a computational tool, or hypothesis, to study the neural basis of language.

Here we asked a more specific question, that is, how a sequence of spoken words is incrementally structured and coherently interpreted in the brain. Our goal was to develop quantitative measures of sentence structure that capture the interplay between different types of constraints that simultaneously influence this process. As a potential solution, we extracted detailed structural measures specific to the contents in each sentence from the hidden states of BERT, which was trained on massive corpora from real-life language use. Although DLMs such as BERT are not specifically designed to parse sentences, they can learn from training corpora the multi-dimensional properties related to sentence structure and dependency[60]. In line with this, our analyses confirmed that BERT structural measures incorporate relevant lexical constraints and that they exhibit both behavioral and neural alignments with human listeners.

Taking advantage of the contextualized BERT structural measures, our RSA results provide neural evidence for the construction of a coherent interpretation driven by the interaction between linguistic and non-linguistic knowledge evoked by individual words as they are heard sequentially in a spoken sentence. Specifically, neural representations of an unfolding sentence's structure initially emerged in bilateral fronto-temporal regions and became left-lateralized when more complex syntactic properties, rather than canonical linear adjacency, were considered to build a structured interpretation (e.g., beyond Verb1 in our stimulus sentences). Meanwhile, we found considerable right-hemisphere effects for computations associated with broad world knowledge, which is essential for understanding the intended meaning conveyed by the speaker[61]. In addition to the core fronto-temporal language network, we found that the multiple-demand network and the default mode network were also involved during online construction of structured interpretations, which may reflect additional cognitive demands for resolving potential structural ambiguity and evaluating the plausibility of underlying events[62].

There are two points to note about the use of BERT. Firstly, unlike autoregressive DLMs trained using left-to-right attention and next-word prediction, BERT is trained to predict masked words in a sentence with a bi-directional attention mechanism. The additional right-to-left attention provides updated representations of preceding words every time an incoming word is added to the input (e.g., representation of *"dog"* in *"The dog…"* is different from that in *"The dog found…"*). This feature of BERT is useful for tracking the dynamic change of the representation of a specific word as its context evolves, particularly in sentences with structural ambiguity. Although autoregressive DLMs also update hidden states as the input unfolds and could be used to study complex sentential structures[63], the updated contextual effects are reflected in the hidden states of the right-most incoming word, while those of the preceding

words on the left remain unchanged (i.e., the representation of *"dog"* is the same in *"The dog…"* and *"The dog found…"*). This is different from BERT, where the updated contextual effects are reflected in the hidden states of all preceding words in both directions.

Secondly, although we input each sentence word-by-word to BERT, however, unlike human listeners or recurrent neural networks, BERT process two consecutive inputs (e.g., *"The dog…"* and *"The dog found…"*) independently in a parallel manner, and there is no direct relationship between these two inputs. In fact, human listeners would not start over from the beginning of a sentence as it unfolds word-by-word, but update it continually as each word is heard. They use whatever information currently available to build a coherent interpretation[64]. Nevertheless, this discrepancy does not hinder our goal of extracting contextualized structural measures from sentence fragments that approximate the current structured interpretation. The representation of each word is continuously updated in a bi-directional way as a new word is added to the input, taking into account the constraints placed by the specific words and their interaction to form a coherent interpretation.

In summary, recent developments in DLMs have shown great potential in capturing the dynamic interplay between syntax, semantics, and world knowledge that is essential for successful language comprehension. As demonstrated in this study, when considered as putative brain-computational models and combined with advanced neuroimaging methods within an appropriate framework[51], future DLMs, with more human-like model architecture[65] and rigorous evaluation[66], may provide new insights into the neural implementation of the various incremental processing operations that support the rapid transition from sound to meaning in the brain.

Materials and Methods

Details of materials and methods are provided in Supplementary Materials.

Acknowledgements

This research was funded by European Research Council Advanced Investigator Grant to L.K.T. under the European Community's Horizon 2020 Research and Innovation Programme (2014-2022 ERC Grant Agreement 669820). B.L. was supported by Changping Laboratory. We thank Billi Randall and Barry Devereux for their valuable contributions to early experimental design and to stimulus development; and Hun S. Choi, Benedict Vassileiou, John Hewitt, Tao Li, Yi Zhu, Nai Ding and Giorgio Marinato for helpful discussions.

Author contributions

Conceptualization: L.K.T., W.D.M., B.L.
Investigation, Data curation: B.L., Y.F.
Methodology, Formal Analysis & Visualization: B.L.
Funding acquisition & Project administration: L.K.T.
Supervision: L.K.T., W.D.M.
Writing – original draft, review & editing: B.L., W.D.M., L.K.T.

Declaration of interests

Authors declare no competing interests.

Data and materials availability

Upon publication, data and code will be made available online.

References

1. Marslen-Wilson, W. & Tyler, L.K. The temporal structure of spoken language understanding. *Cognition* **8**, 1-71 (1980).
2. Choi, H.S., Marslen-Wilson, W.D., Lyu, B., Randall, B. & Tyler, L.K. Decoding the Real-Time Neurobiological Properties of Incremental Semantic Interpretation. *Cereb Cortex* **31**, 233-247 (2021).
3. Lyu, B.*, et al.* Neural dynamics of semantic composition. *Proc Natl Acad Sci U S A* **116**, 21318-21327 (2019).
4. Everaert, M.B.H., Huybregts, M.A.C., Chomsky, N., Berwick, R.C. & Bolhuis, J.J. Structures, Not Strings: Linguistics as Part of the Cognitive Sciences. *Trends Cogn Sci* **19**, 729-743 (2015).
5. Matchin, W. & Hickok, G. The Cortical Organization of Syntax. *Cereb Cortex* **30**, 1481-1498 (2020).
6. Nelson, M.J.*, et al.* Neurophysiological dynamics of phrase-structure building during sentence processing. *Proc Natl Acad Sci U S A* **114**, E3669-E3678 (2017).
7. Law, R. & Pylkkanen, L. Lists with and without Syntax: A New Approach to Measuring the Neural Processing of Syntax. *J Neurosci* **41**, 2186-2196 (2021).
8. Pallier, C., Devauchelle, A.D. & Dehaene, S. Cortical representation of the constituent structure of sentences. *Proc Natl Acad Sci U S A* **108**, 2522-2527 (2011).
9. Friederici, A.D., Bahlmann, J., Heim, S., Schubotz, R.I. & Anwander, A. The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc Natl Acad Sci U S A* **103**, 2458-2463 (2006).
10. Bever, T.G. The cognitive basis for linguistic structures. in *Cognition and the Development of Language* (ed. J.R. Hayes) (John Wiley, New York, 1970).
11. Tyler, L.K. & Marslen-Wilson, W.D. The On-Line Effects of Semantic Context on Syntactic Processing. *J Verbal Learn Verbal Behav* **16**, 683-692 (1977).
12. MacDonald, M.C., Pearlmutter, N.J. & Seidenberg, M.S. The lexical nature of syntactic ambiguity resolution. *Psychol Rev* **101**, 676-703 (1994).
13. Trueswell, J.C. & Tanenhaus, M.K. Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. in *Perspectives on sentence processing.* 155-179 (Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 1994).
14. Altmann, G.T. Ambiguity in sentence processing. *Trends Cogn Sci* **2**, 146-152 (1998).
15. Vaswani, A.*, et al.* Attention is all you need. *Advances in neural information processing systems* **30** (2017).
16. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* 4171-4186 (Association for Computational Linguistics, Minneapolis, MN, USA, 2019).
17. Brown, T.*, et al.* Language models are few-shot learners. *Advances in neural information processing systems* **33**, 1877-1901 (2020).
18. Ouyang, L.*, et al.* Training language models to follow instructions with human feedback. *Advances in neural information processing systems* (2022).
19. Bisk, Y.*, et al.* Experience Grounds Language. in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* 8718-8735 (Association for Computational Linguistics, Online, 2020).
20. Linzen, T. & Baroni, M. Syntactic Structure from Deep Learning. *Annu Rev Linguist* **7**, 195-212 (2020).
21. Pavlick, E. Semantic Structure in Deep Learning. *Annu Rev Linguist* **8**, 447-471 (2022).

22. Mahowald, K.*, et al.* Dissociating language and thought in large language models: a cognitive perspective. *arXiv* (2023).
23. Marslen-Wilson, W.D. Sentence perception as an interactive parallel process. *Science* **189**, 226-228 (1975).
24. Kuperberg, G.R. Neural mechanisms of language comprehension: challenges to syntax. *Brain Res* **1146**, 23-49 (2007).
25. Bengio, Y., Lecun, Y. & Hinton, G. Deep Learning for AI. *Commun Acm* **64**, 58-65 (2021).
26. Elman, J.L. Finding Structure in Time. *Cognitive Sci* **14**, 179-211 (1990).
27. Elman, J.L. Learning and development in neural networks: the importance of starting small. *Cognition* **48**, 71-99 (1993).
28. Schrimpf, M.*, et al.* The neural architecture of language: Integrative modeling converges on predictive processing. *Proc Natl Acad Sci U S A* **118**, e2105646118 (2021).
29. Goldstein, A.*, et al.* Shared computational principles for language processing in humans and deep language models. *Nat Neurosci* **25**, 369-380 (2022).
30. Heilbron, M., Armeni, K., Schoffelen, J.M., Hagoort, P. & de Lange, F.P. A hierarchy of linguistic predictions during natural language comprehension. *Proc Natl Acad Sci U S A* **119**, e2201968119 (2022).
31. Toneva, M., Mitchell, T.M. & Wehbe, L. Combining computational controls with natural text reveals aspects of meaning composition. *Nat Comput Sci* **2**, 745-757 (2022).
32. Caucheteux, C., Gramfort, A. & King, J.R. Deep language algorithms predict semantic comprehension from brain activity. *Sci Rep* **12**, 16327 (2022).
33. Caucheteux, C. & King, J.R. Brains and algorithms partially converge in natural language processing. *Commun Biol* **5**, 134 (2022).
34. Caucheteux, C., Gramfort, A. & King, J.R. Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nat Hum Behav* **7**, 430-441 (2023).
35. Hewitt, J. & Liang, P. Designing and interpreting probes with control tasks. in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* 2733-2743 (Association for Computational Linguistics, Hong Kong, China, 2019).
36. Tenney, I.*, et al.* What do you learn from context? probing for sentence structure in contextualized word representations. in *the 7th International Conference on Learning Representations* (New Orleans, LA, USA, 2019).
37. Hewitt, J. & Manning, C.D. A structural probe for finding syntax in word representations. in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* 4129-4138 (Association for Computational Linguistics, Minneapolis, MN, USA, 2019).
38. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci* **2**, 4 (2008).
39. Dowty, D. Thematic Proto-Roles and Argument Selection. *Language* **67**, 547-619 (1991).
40. Marslen-Wilson, W.D., Tyler, L.K. & Koster, C. Integrative Processes in Utterance Resolution. *J Mem Lang* **32**, 647-666 (1993).
41. Friederici, A.D. The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences* **16**, 262-268 (2012).
42. Yeshurun, Y., Nguyen, M. & Hasson, U. The default mode network: where the idiosyncratic self meets the shared social world. *Nat Rev Neurosci* **22**, 181-192 (2021).
43. De Marneffe, M.-C., MacCartney, B. & Manning, C.D. Generating typed dependency parses from phrase structure parses. in *Proceedings of the 5th International Conference on*

Language Resources and Evaluation 449-454 (European Language Resources Association, Genoa, Italy, 2006).
44. Jackendoff, R. & Jackendoff, R.S. *Foundations of language: Brain, meaning, grammar, evolution* (Oxford University Press, USA, 2002).
45. Duncan, J. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn Sci* **14**, 172-179 (2010).
46. Baldassano, C.*, et al.* Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* **95**, 709-721 e705 (2017).
47. Humphreys, G.F., Lambon Ralph, M.A. & Simons, J.S. A Unifying Account of Angular Gyrus Contributions to Episodic and Semantic Cognition. *Trends Neurosci* **44**, 452-463 (2021).
48. Jung-Beeman, M. Bilateral brain processes for comprehending natural language. *Trends Cogn Sci* **9**, 512-518 (2005).
49. Metusalem, R., Kutas, M., Urbach, T.P. & Elman, J.L. Hemispheric asymmetry in event knowledge activation during incremental language comprehension: A visual half-field ERP study. *Neuropsychologia* **84**, 252-271 (2016).
50. Troyer, M., McRae, K. & Kutas, M. Wrong or right? Brain potentials reveal hemispheric asymmetries to semantic relations during word-by-word sentence reading as a function of (fictional) knowledge. *Neuropsychologia* **170**, 108215 (2022).
51. Kriegeskorte, N. & Douglas, P.K. Cognitive computational neuroscience. *Nat Neurosci* **21**, 1148-1160 (2018).
52. Yamins, D.L. & DiCarlo, J.J. Using goal-driven deep learning models to understand sensory cortex. *Nat Neurosci* **19**, 356-365 (2016).
53. Rabovsky, M., Hansen, S.S. & McClelland, J.L. Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nat Hum Behav* **2**, 693-705 (2018).
54. Donhauser, P.W. & Baillet, S. Two Distinct Neural Timescales for Predictive Speech Processing. *Neuron* (2019).
55. Yang, G.R., Joglekar, M.R., Song, H.F., Newsome, W.T. & Wang, X.J. Task representations in neural networks trained to perform many cognitive tasks. *Nat Neurosci* **22**, 297-306 (2019).
56. Kietzmann, T.C.*, et al.* Recurrence is required to capture the representational dynamics of the human visual system. *Proc Natl Acad Sci U S A* **116**, 21854-21863 (2019).
57. Bao, P., She, L., McGill, M. & Tsao, D.Y. A map of object space in primate inferotemporal cortex. *Nature* **583**, 103-108 (2020).
58. Sheahan, H., Luyckx, F., Nelli, S., Teupe, C. & Summerfield, C. Neural state space alignment for magnitude generalization in humans and recurrent networks. *Neuron* **109**, 1214-1226 e1218 (2021).
59. Giordano, B.L., Esposito, M., Valente, G. & Formisano, E. Intermediate acoustic-to-semantic representations link behavioral and neural responses to natural sounds. *Nat Neurosci* (2023).
60. Manning, C.D., Clark, K., Hewitt, J., Khandelwal, U. & Levy, O. Emergent linguistic structure in artificial neural networks trained by self-supervision. *Proc Natl Acad Sci U S A* **117**, 30046-30054 (2020).
61. Bicknell, K., Elman, J.L., Hare, M., McRae, K. & Kutas, M. Effects of event knowledge in processing verbal arguments. *J Mem Lang* **63**, 489-505 (2010).
62. Smallwood, J.*, et al.* The default mode network in cognition: a topographical perspective. *Nature Reviews Neuroscience* **22**, 503-513 (2021).
63. Jurayj, W., Rudman, W. & Eickhof, C. Garden Path Traversal in GPT-2. in *Proceedings of the 5th BlackboxNLP Workshop on Analyzing and Interpreting Neural*

Networks for NLP 305-313 (Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022).
64. Frazier, L. & Rayner, K. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cogn Psychol* **14**, 178-210 (1982).
65. McClelland, J.L., Hill, F., Rudolph, M., Baldridge, J. & Schutze, H. Placing language in an integrated understanding system: Next steps toward human-level performance in neural language models. *Proc Natl Acad Sci U S A* **117**, 25966-25974 (2020).
66. Binz, M. & Schulz, E. Using cognitive psychology to understand GPT-3. *Proc Natl Acad Sci U S A* **120**, e2218523120 (2023).