Traces in the Brain: Neural Evidence for Syntactic Movement in English and Chinese

Yuhan Huang[1], Zhengwu Ma[1], Yuqi Jin[1], Beth Chan[2], Zheng Shen[2], Jackie Yan-Ki Lai[1], John T. Hale[3], Jixing Li[1]*

[1]Department of Linguistics and Translation, City University of Hong Kong,
[2]Department of English, Linguistics and Theatre Studies, National University of Singapore,
[3]Department of Cognitive Science, Johns Hopkins University,

{yuhan.huang, zhengwu.ma, yuqijin5}@my.cityu.edu.hk, bethchf@u.nus.edu,
zhengshen@nus.edu.sg, {jyk.lai, jixingli}@cityu.edu.hk, jthale@jhu.edu

Abstract

Syntactic movement is a core concept in generative linguistics to account for word-order variation and long-distance dependencies, but its psychological and neurobiological status remains debated. Here, we test the neural reality of movement in English and Chinese by correlating brain activity during naturalistic listening with syntactic node counts, traces and word embeddings derived from X-bar style tree annotations. We find that deep structure significantly predicts neural responses in English but not in Chinese, providing partial support for movement-based accounts while revealing clear cross-linguistic differences.

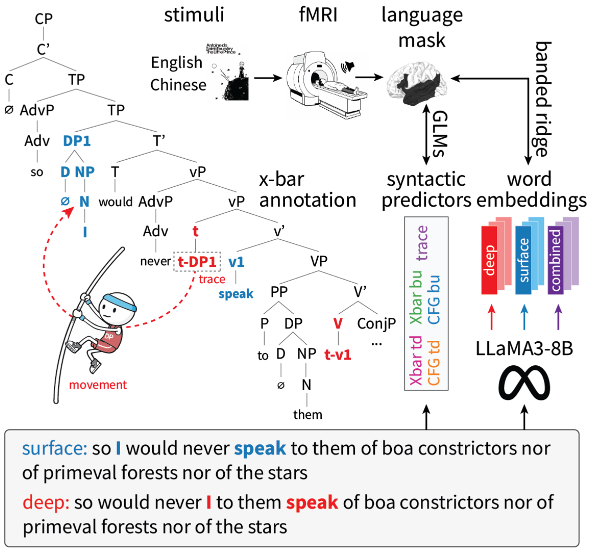

Figure 1: Aligning X-bar–based syntactic predictors and word embeddings with fMRI data during naturalistic listening in English and Chinese.

1 Introduction

Syntactic movement refers to the displacement of constituents from their base positions to derive surface word order, a core idea in generative grammar and X-bar theory (Chomsky, 1970; Jackendoff, 1977; Chomsky, 1981). These frameworks distinguish an underlying deep structure, where thematic roles are assigned, from a surface structure that determines linear order. Movement preserves hierarchical structure while producing non-canonical word orders and leaving a trace at the original position. For example, in passivization, "John hit Bob" is derived as "Bob was hit *t* (by John)", where "Bob" moves from object to subject position, leaving a trace *t* in object position.

Despite its formal elegance, movement-based syntax has been challenged by accounts that explain non-canonical word order without invoking transformations or multiple representational levels. Usage-based and constructionist approaches question the need for deep–surface representations, citing the lack of clear neurobiological evidence. These critiques are especially salient cross-linguistically, as movement-based analyses developed for English do not straightforwardly extend to languages such as Chinese, where dependencies are often resolved *in situ* or through discourse-pragmatic mechanisms.

To empirically evaluate the neurobiological evidence for syntactic movement, we manually annotated every sentence in the audiobook *The Little Prince* in both English (N=1502) and Chinese (N=1577) using X-bar–style syntactic representations. Based on these annotations, we derived syntactic node counts under both top-down and bottom-up parsing strategies and correlated these measures with brain activity during naturalistic listening. To capture the hypothesized processing cost of movement, we additionally coded a binary regressor at trace positions, aligned to the offset of the preceding overt word in the dependency, to index the processing of an unpronounced copy of the moved element. In parallel, we extracted contextual word embeddings from each layer of LLaMA3.1-8B (Grattafiori et al., 2024) and constructed representations reflecting both deep and

*Corresponding author.

surface word order. These embeddings were then correlated with fMRI responses while English and Chinese participants listened to the audiobook, using an openly available naturalistic fMRI dataset (Li et al., 2022). Figure 1 provides an overview of the analysis pipeline. Our findings reveal that:

- Syntactic node counts derived from X-bar annotations show stronger correlations with brain activity than simpler Penn Treebank–style context-free grammar (CFG) annotations in the left temporal regions for both English and Chinese, with additional engagement of left frontal regions for English under top-down parsing;
- Trace-based regressors significantly predict neural activity in the left temporal regions for both languages, and additionally in left frontal regions for English;
- LLaMA-derived word embeddings constructed in deep-structure word order predict additional activity in the left superior temporal gyrus relative to surface-order embeddings in English, whereas no contrast is observed in Chinese.

Taken together, these results provide partial neurobiological support for syntactic movement and a distinction between deep and surface structural representations during online sentence processing in English, but not in Chinese, suggesting that the neural relevance of syntactic movement may be language-specific rather than a universal property of human sentence processing.

2 Related Work

2.1 X-bar theory and syntactic movement

X-bar theory developed in the 1970s by Chomsky, Jackendoff, and others (Chomsky, 1970; Jackendoff, 1977) posits that all phrases share a similar three-level structure: a head X (lexical level), which projects to an intermediate X' level (head plus complement), and a full phrase XP (possibly with a specifier). This schema applies to all categories (DP, VP, PP, CP, etc.). Crucially, X-bar theory provides landing sites for moved constituents (typically specifier or adjunct positions) and constrains movement so that it does not violate the structural template. Movement in X-bar theory is typically characterized as an *upward movement* to a *c-commanding* position that is initially empty. In practice, this means an element can only move to a higher position in the tree, often a specifier or head position above its starting point. A *trace*, notated as t or sometimes *ec*, is left behind in the original site. The movement and trace together form a chain: the moved element and its trace share a single interpretation. Syntactic constraints like the Proper Binding Condition and Empty Category Principle (ECP) were proposed to regulate these chains, requiring for instance that certain types of traces be properly governed or bound (Fiengo, 1977; Chomsky, 1981). In essence, traces are null categories that carry grammatical features but no phonetic content, ensuring that despite the changed word order, the sentence can be interpreted as if the displaced element were in its base position.

2.2 Movement in English and Chinese

Generative syntax classifies movements by what category moves and where it moves to. The major types include A-movement, A'-movement and head movement. A-movement refers to argument movement to *Case* position, typically DP/NP movement to subject or object positions. Classic instances of DP-movement include passivization and raising constructions. In a passive sentence, the object of an active verb becomes the surface subject. For example, "John$_1$ hit Bob$_2$." becomes "Bob$_2$ was hit t_2 (by John$_1$)." Here the DP "Bob$_2$" originates as the object of the verb "hit" but moves to the subject position, leaving a trace t_2 as the verb's object. The trace carries the index of "Bob," so the verb "hit" can still assign its theme role to that position. English marks this movement with auxiliary "be" + past participle. In Chinese, passivization is marked by the "被*(bei)*-construction" as in "张三$_1$被李四$_2$打了t_1." ("Zhangsan$_1$ was hit t_1 by Lisi$_2$"). Early generative analyses debated whether *bei*-passives involve NP-movement like English (Wang, 1970) or are base-generated as a different structure (e.g., Hashimoto, 1969, 1987), illustrating the debate of whether Chinese employs the same kind of movement as English.

A'-movement refers to non-argument movement to operator positions, such as *wh*-phrases to clause-initial position. In English, A'-movement is overt and obligatory in *wh*-questions: for example, in "What did Mary buy t?" the object what moves from its base position as the complement of "buy" to the clause-initial Spec-CP position, leaving a trace that marks its thematic role. In contrast, Chinese is a *wh-in-situ* language and does not exhibit overt A'-movement of *wh*-phrases. In a sentence such as "张三买了什么？" ("Zhangsan bought what?"), the *wh*-object "什么 (what)" remains in its canonical postverbal position, and no surface displacement occurs.

A'-movement also includes movement of adverbial elements (AdvPs) like adjunct fronting. In English, adjuncts such as adverbs or prepositional phrases can be fronted for emphasis, topic, or focus. For instance: "Quickly, John finished the report *t*." English also allows fronting of negative adverbs trigger inversion, such as "Never have I *t* seen that," or "Only then did he realize *t* his mistake." In Chinese, adverbial phrases in general do not need to move to achieve various discourse functions because the base-generated order is already quite flexible for adjuncts in Chinese. Temporal and locative adverbials typically appear at the beginning of the sentence in Chinese by default. For example, in the sentence "昨天张三在这里看书。" ("Yesterday Zhangsan read a book here."), the AdvP "昨天 (yesterday)" is at the front without any marked inversion.

Head movement involves a head (like a verb, Tense, or Complementizer) moving to a higher head position. A prototypical case is T-to-C movement in English questions. In an English *yes-no* question or *wh*-question, an auxiliary or modal in T (the inflectional head) moves to the C (complementizer) position at the front of the clause. For example, "John has eaten." becomes "Has John eaten?", where the T head "has" moves to a higher position in the CP domain. If no auxiliary or modal is present in a main clause (e.g. "John saw Mary."), English syntax supplies the dummy "do" to support T-to-C movement ("Did John see Mary?"), a process known as *do*-support, which indirectly reflects that T must move to C in questions even when T is just a tense affix. In Chinese, T or inflectional heads do not undergo analogous movement. Chinese lacks auxiliary inversion in questions; instead, it employs other strategies: a sentence-final question particle "吗 (ma)" for yes-no questions, or the A-not-A construction such as "他去不去?" ("He go-not-go?") to mean "Is he going or not?" for yes-no questions without particles.

The absence of overt *wh*-movement and head movement in yes/no questions in Chinese does not entail the absence of corresponding functional projections (e.g., TopicP, CP, FocusP). In many generative accounts, *in situ* elements in Chinese are assumed to undergo covert movement to the same structural positions as in English, with displacement occurring at Logical Form rather than being phonologically realized (Huang, 1982).

In sum, English and Chinese differ systematically in how syntactic dependencies are expressed, with English relying more heavily on overt syntactic movement and Chinese exhibiting comparatively less overt displacement. While Chinese does allow topic fronting and DP movement in *bei*-passives, other constructions like yes-no questions and *wh*-questions are not expressed by overt movement. Consequently, syntactic dependencies in Chinese are more often resolved *in situ* or through pragmatic mechanisms, highlighting a cross-linguistic asymmetry in the relationship between deep and surface structural representations.

2.3 Debates on syntactic movement

From the inception of transformational syntax, Chomskyan formalism has posited distinct levels of representation (deep structure vs. surface structure) to explain how one underlying proposition can correspond to multiple surface orders (Chomsky, 1955, 1957, 1965, 1970, 1981; Jackendoff, 1977). *Movement* serves as the mechanism linking these representational levels. Its primary motivation was explanatory unification: movement allows superficially distinct constructions to share a common underlying structure, avoiding construction-specific rules. For example, actives and passives are analyzed as sharing a common underlying argument structure, with passivization involving displacement of the object to subject position. Likewise, interrogatives are derived from declaratives through operator movement—generalizations that were difficult to capture in the non-transformational grammars of the 1950s.

However, over the years many critiques have arisen. One line of critique argued that a lot of what generative grammar attributes to transformations and abstract structure can be explained by resource limitations and parsing strategies. For example, the Dependency Locality Theory (DLT; Gibson, 1998, 2000) posits that the difficulty of a sentence largely comes from the distance between dependent elements and the number of new referents introduced between them. This is a gradient notion, not a strict grammatical constraint. It does not distinguish "deep" vs. "surface" structures, there is just the surface and the memory cost to connect elements in it. This approach essentially downplays the necessity of an independent "deep structure",

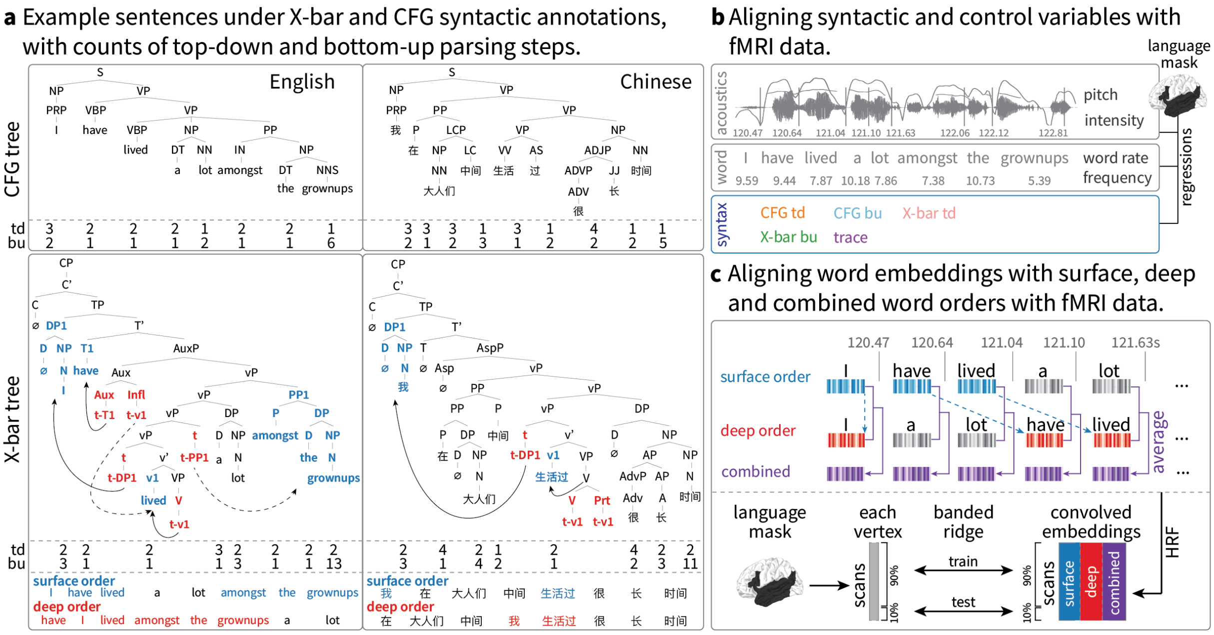

Figure 2: Methods overview. **a** Example English and Chinese sentences under X-bar–style and CFG syntactic annotations, with counts of top-down and bottom-up parsing steps. **b** Aligning syntactic and control variables with fMRI data using GLM. **c** Aligning LLaMA3 embeddings from surface and deep structures, and their average with fMRI data using banded ridge regression.

as everything that matters is in the surface dependency configuration and our memory handling of it.

The critique can also be framed in terms of representational economy: in the absence of clear neurobiological evidence for multiple syntactic representations, usage-based approaches (Abbot-Smith and Tomasello, 2010; Ambridge and Lieven, 2011; Arnon and Snider, 2010; Boas, 2008; Diessel and Hilpert, 2016; Dunn, 2019; Kidd et al., 2010; Ibbotson, 2022; Goldberg, 1995, 2006) favor a single level of representation in which structure emerges from distributional regularities, dispensing with movement in favor of direct form–meaning pairings. Nonetheless, psycholinguistic and neurolinguistic findings provide support for movement-based analyses. Large-scale acceptability studies show sharp island effects not fully explained by gradient memory metrics (Sprouse et al., 2012), and neuroimaging work demonstrates that *wh*-movement and head movement engage distinct brain regions, with only the former overlapping canonical syntactic networks (Shetreet and Friedmann, 2014). These results suggest that the brain differentiates movement types in ways consistent with their formal linguistic analyses.

Building on this line of work, the present study aims to directly test the neurobiological reality of syntactic movement and representational levels under naturalistic language comprehension. Rather than focusing on isolated sentence types or controlled contrasts, we combine fine-grained X-bar–style syntactic annotations with fMRI data collected during audiobook listening in both English and Chinese. This design allows us to ask whether neural responses track movement-related representations such as traces and deep-structure word order beyond surface word order and general processing demands, and whether such effects generalize across typologically distinct languages.

3 Methods

3.1 Annotations

X-bar trees. The English version of the audiobook *The Little Prince* comprises 15,369 words across 1,502 sentences, with a mean sentence length of 10.23 words (SD = 6.92). The Chinese version comprises 15,195 words across 1,577 sentences, with a mean sentence length of 9.64 words (SD = 6.00). Two trained syntacticians (one native English speaker and one native Mandarin Chinese speaker) manually annotated X-bar trees for each sentence in the English and Chinese. All annotations followed a unified X-bar schema, with additional Chinese-specific functional heads where appropriate (see Table 2 in Appendix A for details).

Language	Word order preserved	Word order changed	Word order changed ratio
English	645	857	57.1%
Chinese	363	1214	77.0%

Table 1: Proportion of sentences exhibiting word order changes induced by movement operations in English and Chinese stimuli.

Annotation was carried out independently and resolved through discussion to ensure consistency.

We identified 857 English sentences and 1,214 Chinese sentences in which deep-structure word order differed from surface word order, corresponding to word-order change ratios of 57.1% and 77.0%, respectively (see Table 1). This result appears counterintuitive, given that Chinese is typically characterized as involving less overt movement than English. Closer inspection reveals that movement in Chinese more frequently results in word-order changes than English. For example, in the English sentence "Here is a copy of the drawing", the clause involves DP movement to sentence-initial position, but the moved DP is a null category (∅), yielding no visible change in surface word order. In contrast, in the Chinese sentence "页头上就是那幅画的摹本。" ("Page-top on exactly is that classifier (CL) painting 's copy".), the PP "页头上 (page-top on)" is overtly fronted to sentence-initial position, producing a clear surface word-order change (see Figure 6 in Appendix A)

Traces. Within this annotation scheme, traces were represented as terminal nodes labeled with *t* plus their syntactic category and index (e.g., t-DP$_i$, t-V$_i$). This procedure yielded 5,169 traces in the English sentences and 3,435 traces in the Chinese sentences. Table 3 in Appendix A summarizes the counts of different trace types in the English and Chinese texts. Based on these annotations, we derived a binary *trace* regressor that was set to 1 at the offset of the word immediately preceding the trace, and to 0 at all other positions.

CFG trees. For comparison, we also generated Penn Treebank–style context-free grammar (CFG) trees for all sentences using Stanford CoreNLP (v4.5.10; Manning et al., 2014).

Node count. Node count at a given word was defined as the number of parser actions required to advance from that word to the next under a specified parsing strategy. In top-down parsing, the parser projects hierarchical structure from higher-level nodes before matching the input string, whereas in bottom-up parsing, the parser incrementally builds structure from terminal words upward, applying phrase-structure rules only after lexical evidence is available (Hale, 2014). We computed node counts for each word based on both X-bar and CFG trees under each parsing strategy. Figure 2a illustrates differences in node counts derived from X-bar and CFG trees under different parsing strategies.

Controlled variables. We included four control variables: pitch, intensity, word rate, and word frequency, which are known to correlate with activity in core language-related brain regions and served as baseline regressors against which syntactic effects were evaluated. **Pitch** and root mean square **intensity** were computed at 10-ms resolution for the English and Chinese audiobook using the Voicebox toolbox [1]. **Word rate** was modeled as a binary regressor marking the offset of each word in the audiobook. **Word frequency** was defined as the log-transformed unigram frequency of each word, estimated from Google Ngrams (version 2012070129).

3.2 fMRI data

We analyzed the English and Chinese subsets of a publicly available fMRI dataset (Li et al., 2022), comprising 49 native English speakers (30 females; 21.3 ± 3.6 years) and 35 native Chinese speakers (15 females; 19.3 ± 1.6 years). Participants listened to language-matched audiobooks of *The Little Prince* during scanning (English: 94 min; Chinese: 99 min) in a single session consisting of nine runs of approximately 10 minutes each. Details of fMRI data acquisition and preprocessing are provided in Appendix B.

3.3 Aligning syntactic and control variables with fMRI data

The nine convolved regressors were fit to each participant's fMRI time series at each cortical vertex within a left-lateralized language mask. The mask (black region in Figure 2b,c) included the left temporal lobe, the left inferior frontal gyrus (LIFG; Brodmann areas 44 and 45), and the left angular gyrus (LAG; BA39). The same regression procedure was applied independently to the English and Chinese groups.

[1] http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

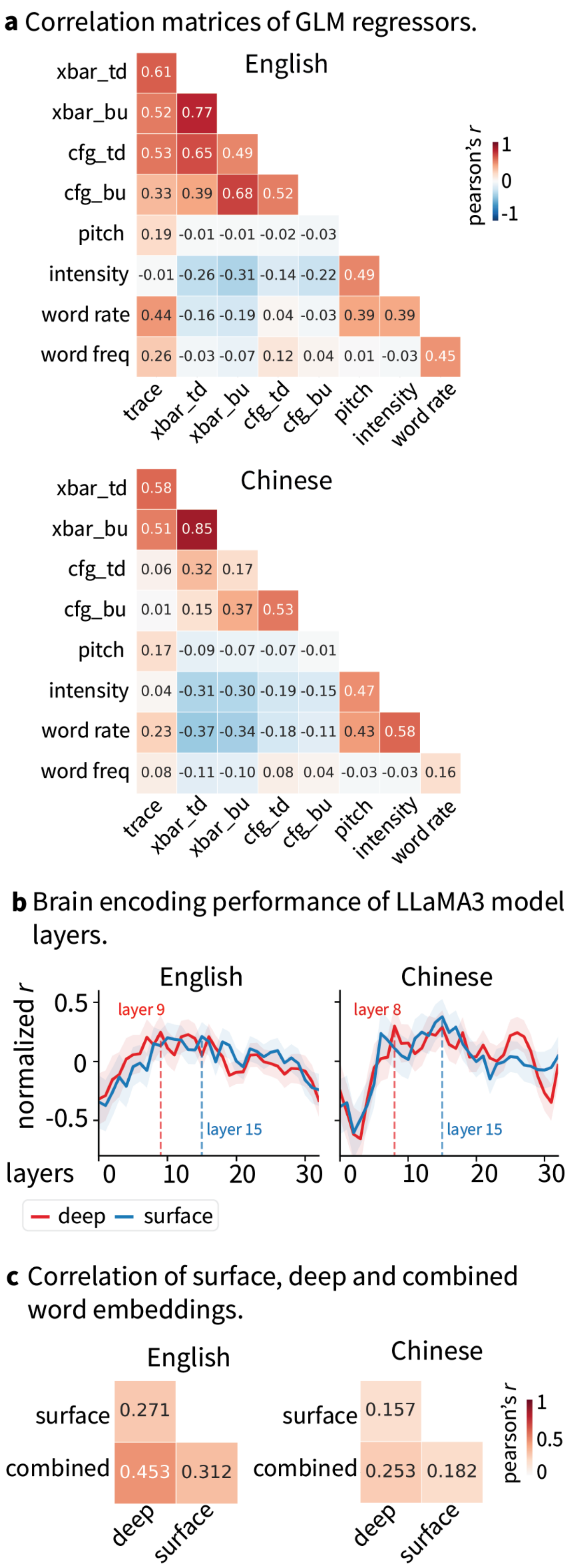

Figure 3: **a** Correlation matrices of syntactic and control regressors for English and Chinese. **b** Layer-wise encoding performance of LLaMA3 against fMRI data in English and Chinese. **c** Correlation matrices of surface-, deep-, and combined-order word embeddings from their best-performing LLaMA3 layers.

3.4 Aligning deep and surface-order word embeddings with fMRI data.

We extracted deep- and surface-structure word embeddings from each layer of a large language model and conducted layer-wise encoding analyses within a predefined language mask. We used the widely adopted open-source LLaMA3.1-8B (Grattafiori et al., 2024), which supports both English and Chinese and has been extensively used in model–brain alignment research. We operationalized "deep structure" as a reordering of identical lexical items based on X-bar dependencies, and aligned word embeddings following this deep-structure order to the same temporal offsets as their surface-order counterparts in the audio (see Figure 2c). For each language, ridge regression models were trained to predict neural responses at each cortical vertex using embeddings from eight data segments and tested on the held-out ninth segment. To identify the best-performing layer, we averaged Pearson's r across vertices within the language mask, yielding a single encoding score per layer.

We next modeled neural responses using deep- and surface-order word embeddings from the best-performing LLaMA3 layer with banded ridge (multi-kernel) regression (Dupré la Tour et al., 2022), implemented via Himalaya's MultipleKernelRidgeCV. In addition to the individual deep and surface embeddings, we included a combined embedding defined as their average (see Appendix C for details of the banded ridge regression). Prior to regression, all embeddings were reduced to their first 100 principal components. Each embedding set was treated as a separate kernel with its own regularization parameter, enabling joint integration of multiple embedding spaces while adaptively weighting their contributions.

3.5 Statistical significance test

At the group level, regression coefficients (β) from the GLM and Pearson's r values from the banded ridge regression were first z-scored and evaluated separately for each language group using one-sample, one-tailed t-tests with a cluster-based permutation procedure (Maris and Oostenveld, 2007). Statistical significance was assessed using 10,000 permutations. Clusters were defined by an initial threshold of $p < 0.05$ and were required to span at least 20 vertices. All group-level analyses were performed using the Python packages MNE (v1.10.2; Gramfort et al., 2014) and Eelbrain (v0.41; Brod-

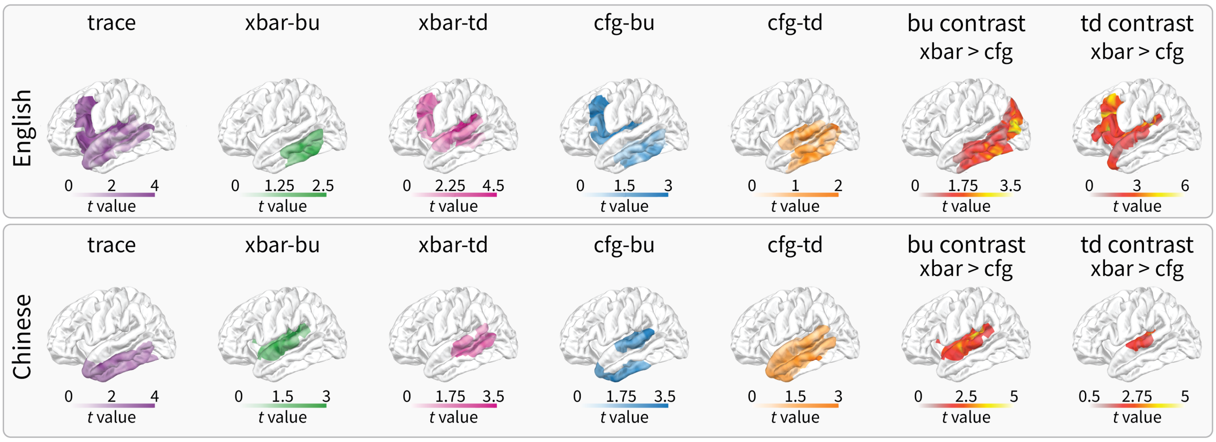

Figure 4: Significant clusters associated with trace and syntactic node counts in English and Chinese.

beck et al., 2023).

4 Results

4.1 Correlation among regressors

The nine regressors, including X-bar top-down and bottom-up node counts (xbar-td, xbar-bu), CFG top-down and bottom-up node counts (cfg-td, cfg-bu), trace, pitch, intensity, word rate, and word frequency, were first convolved with the canonical hemodynamic response function (HRF) to match the temporal resolution of the fMRI data. We then computed correlation matrices among these convolved regressors. The strongest correlation was observed between xbar-bu and xbar-td in both English ($r = 0.77$) and Chinese ($r = 0.85$), whereas no other regressor pair exceeded a correlation of 0.7 (see Figure 3a).

4.2 Encoding performance across LLaMA3 model layers

The best-performing layers for deep- and surface-structure embeddings were layer 9 and layer 15 for English, and layer 8 and layer 15 for Chinese, respectively (see Figure 3b). Correlations between surface-, deep-, and combined-order word embeddings from their best-performing layers showed highest correlation between deep and combined word embeddings in English (r=0.45) and Chinese (r=0.25).

4.3 Brain regions associated with syntactic node counts and traces

We observed significant brain clusters associated with trace and all syntactic node-count regressors in canonical language regions for both English and Chinese. However, node counts derived from X-bar trees, under both top-down and bottom-up parsing, predicted additional neural activity compared to simpler CFG-based measures. Moreover, in English, trace and the contrast between xbar-td and cfg-td elicited additional activation in left frontal regions, suggesting greater sensitivity to syntactic processing in English than in Chinese (see Figure 4 and Table 4 in Appendix D for detailed statistics of the brain clusters). Regression results from control variables are shown in Figure 7 and Table 5 in Appendix D.

4.4 Brain regions associated with deep and surface word embeddings

For English, we observed significant left temporal activation associated with word embeddings extracted from LLaMA3 under both deep- and surface-structure word orders, as well as with their average. Contrast analyses revealed stronger responses for surface-order than deep-order embeddings, while the combined embeddings elicited greater activity than either surface or deep embeddings alone (see the left panel in Figure 5 and Table 6 in Appendix E for detailed statistics).

For Chinese, significant left temporal activation was observed only for surface-order and combined embeddings, with a significant contrast favoring surface-order over deep-order embeddings in this region. The combined embedding did not elicit additional activity relative to the surface-order embedding, and only a small cluster in the superior temporal region was observed for the combined vs. deep contrast (see the right panel in Figure 5 and Table 6 in Appendix E for detailed statistics).

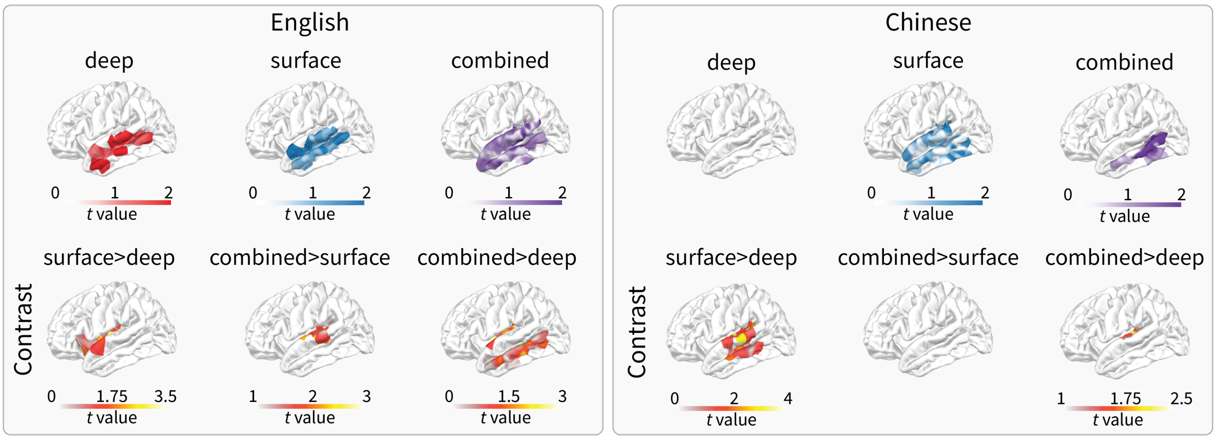

Figure 5: Significant clusters associated with deep-structure, surface-structure and combined word embeddings in English and Chinese.

5 Discussion

Across both English and Chinese, syntactic node counts and trace regressors significantly predicted activity in canonical language regions, particularly in the left temporal cortex, with X-bar–based measures consistently outperforming simpler CFG-based controls. These findings align with prior work linking temporal regions to hierarchical structure building and dependency processing during naturalistic language comprehension (Brennan et al., 2012, 2016; Coopmans et al., 2025; Nelson et al., 2017).

However, clear cross-linguistic differences emerged. Relative to X-bar-based predictors, CFG-based predictors appear to recruit additional frontal activity under bottom-up parsing in English and additional temporal activity in Chinese under top-down parsing. Stronger top-down effects are taken to reflect predictive structure-building mechanisms, whereas stronger bottom-up effects are interpreted as evidence for input-driven integrative computations. The cross-linguistic differences between English and Chinese suggest that the relative balance between predictive and integrative processes may be shaped by language-specific structural properties, such as the extent to which overt movement and hierarchical reordering are systematically exploited in the grammar.

In English, trace effects and X-bar–specific contrasts additionally engaged left inferior frontal regions, and deep-structure embeddings explained neural variance beyond surface-order embeddings, with combined representations yielding the strongest encoding performance. In contrast, Chinese showed robust effects only for surface-order embeddings, with minimal contributions from deep-structure representations and little frontal involvement. This asymmetry aligns with typological differences between the languages, as many syntactic dependencies in Chinese are resolved *in situ* or through discourse-pragmatic mechanisms rather than overt displacement. Together, these findings support a hybrid view in which movement-related representations are neurally instantiated when systematically exploited by a language's grammar, but play a reduced role when surface order and discourse cues suffice.

More broadly, our results align with prior model–brain alignment work showing that neural responses during naturalistic comprehension are well predicted by LLM-derived word embeddings (e.g., Gao et al., 2024, 2025; Goldstein et al., 2022; Huth et al., 2016; Toneva et al., 2022; Schrimpf et al., 2021). The superior performance of combined embeddings, especially in English, suggests that surface-order representations alone are insufficient, and that jointly encoding linear context and hierarchical structure yields better alignment with brain activity.

6 Conclusion

In this work, we conducted a naturalistic, cross-linguistic test of syntactic movement by aligning X-bar–based syntactic predictors and LLM embeddings with fMRI data during audiobook listening. Deep-structure and trace-related effects were substantially stronger in English than in Chinese, providing partial neurobiological support for movement-based theories while highlighting their language-specific neural relevance.

Limitations

Several limitations should be acknowledged. First, deep-structure word order was derived from X-bar annotations, which necessarily reflect theoretical assumptions. Although we mitigate this by comparing against CFG-based controls and by using independent LLM representations, alternative syntactic formalisms could yield different deep representations. Extending this approach to other syntactic theories would further strengthen the conclusions.

Second, although naturalistic stimuli offer ecological validity, they reduce experimental control over specific syntactic contrasts. As a result, movement types and dependency configurations are unevenly distributed, particularly across languages. Future work using targeted manipulations embedded within naturalistic contexts could help isolate specific movement operations more precisely.

Finally, while LLaMA3 provides a powerful multilingual embedding space, it is not a model of human sentence processing per se. Differences between model architectures, training data, or tokenization strategies may influence which representational levels best align with brain activity. Future work comparing multiple model families may help disentangle theory-specific from model-specific effects.

Ethics Statement

The authors declare no competing interests. The X-bar syntactic trees were manually annotated by 2 trained human annotators following the instruction to "annotate each sentence using X-bar–style syntactic representations." Annotators were recruited through advertisements and were compensated for their work in accordance with local rates; they reported no conflicts of interest. The annotation task involved no risks.

The fMRI dataset used in the analysis is publicly available and does not contain sensitive content, such as personal information. The adaptation and use of the fMRI dataset are conducted in accordance with its license.

The model states of LLaMA3.1-8B are utilized solely for research purposes, aligning with its intended use.

Acknowledgments

We thank Yue Fu for helping visualize the syntactic trees.

References

Kirsten Abbot-Smith and Michael Tomasello. 2010. The influence of frequency and semantic similarity on how children learn grammar. *First Language*, 30(1):79–101.

Ben Ambridge and Elena V. M. Lieven. 2011. *Child language acquisition: contrasting theoretical approaches*. Cambridge University Press.

Inbal Arnon and Neal Snider. 2010. More than words: frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1):67–82.

Hans C. Boas. 2008. Determining the structure of lexical entries and grammatical constructions in construction grammar. *Annual Review of Cognitive Linguistics*, 6(1):113–144.

Jonathan Brennan, Yuval Nir, Uri Hasson, Rafael Malach, David J. Heeger, and Liina Pylkkänen. 2012. Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, 120(2):163–173.

Jonathan R. Brennan, Edward P. Stabler, Sarah E. Van Wagenen, Wen-Ming Luh, and John T. Hale. 2016. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, 157-158:81–94.

Christian Brodbeck, Proloy Das, Marlies Gillis, Joshua P Kulasingham, Shohini Bhattasali, Phoebe Gaston, Philip Resnik, and Jonathan Z Simon. 2023. Eelbrain, a Python toolkit for time-continuous analysis with temporal response functions. *eLife*, 12:e85012.

Noam Chomsky. 1955. *The logical structure of linguistic theory*. Plenum.

Noam Chomsky. 1957. *Syntactic structures*. De Gruyter.

Noam Chomsky. 1965. *Aspects of the theory of syntax*. MIT press.

Noam Chomsky. 1970. Remarks on nominalization. *Readings in English transformational grammar*, pages 184–221.

Noam Chomsky. 1981. *Lectures on government and binding*. Foris Publications.

Cas W. Coopmans, Helen de Hoop, Filiz Tezcan, Peter Hagoort, and Andrea E. Martin. 2025. Language-specific neural dynamics extend syntax into the time domain. *PLOS Biology*, 23(1):e3002968.

Holger Diessel and Martin Hilpert. 2016. Frequency effects in grammar. In *Oxford research encyclopedia of linguistics*. Oxford University Press.

Jonathan Dunn. 2019. Frequency vs. Association for constraint selection in usage-based construction grammar. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 117–128. Association for Computational Linguistics.

Tom Dupré la Tour, Michael Eickenberg, Anwar O. Nunez-Elizalde, and Jack L. Gallant. 2022. Feature-space selection with banded ridge regression. *NeuroImage*, 264:119728.

Oscar Esteban, Rastko Ciric, Karolina Finc, Ross W. Blair, Christopher J. Markiewicz, Craig A. Moodie, James D. Kent, Mathias Goncalves, Elizabeth DuPre, Daniel E. P. Gomez, Zhifang Ye, Taylor Salo, Romain Valabregue, Inge K. Amlien, Franziskus Liem, Nir Jacoby, Hrvoje Stojić, Matthew Cieslak, Sebastian Urchs, Yaroslav O. Halchenko, Satrajit S. Ghosh, Alejandro De La Vega, Tal Yarkoni, Jessey Wright, William H. Thompson, Russell A. Poldrack, and Krzysztof J. Gorgolewski. 2020. Analysis of task-based functional MRI data preprocessed with fMRIPrep. *Nature Protocols*, 15(7):2186–2202.

Robert Fiengo. 1977. On trace theory. *Linguistic Inquiry*, 8(1):35–61.

Changjiang Gao, Jixing Li, Jiajun Chen, and Shujian Huang. 2024. Measuring meaning composition in the human brain with composition scores from large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11295–11308. Association for Computational Linguistics.

Changjiang Gao, Zhengwu Ma, Jiajun Chen, Ping Li, Shujian Huang, and Jixing Li. 2025. Increasing alignment of large language models with language processing in the human brain. *Nature Computational Science*, 5(11):1080–1090.

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

Edward Gibson. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita, and Wayne O'Neil, editors, *Image, language, brain*, pages 95–126. MIT Press, Cambridge, MA.

Adele E. Goldberg. 1995. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press, Chicago, IL.

Adele E. Goldberg. 2006. *Constructions at work: The nature of generalization in language*. Oxford University Press, Oxford.

Ariel Goldstein, Zaid Zada, Eliav Buchnik, Mariano Schain, Amy Price, Bobbi Aubrey, Samuel A. Nastase, Amir Feder, Dotan Emanuel, Alon Cohen, Aren Jansen, Harshvardhan Gazula, Gina Choe, Aditi Rao, Catherine Kim, Colton Casto, Lora Fanda, Werner Doyle, Daniel Friedman, Patricia Dugan, Lucia Melloni, Roi Reichart, Sasha Devore, Adeen Flinker, Liat Hasenfratz, Omer Levy, Avinatan Hassidim, Michael Brenner, Yossi Matias, Kenneth A. Norman, Orrin Devinsky, and Uri Hasson. 2022. Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, 25(3):369–380.

Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S. Hämäläinen. 2014. MNE software for processing MEG and EEG data. *NeuroImage*, 86:446–460.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, Danny Wyatt, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Francisco Guzmán, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Govind Thattai, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jack Zhang, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Karthik Prasad, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Kushal Lakhotia, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Maria Tsimpoukelli, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Ning Zhang, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohan Maheswari, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye

Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vítor Albiero, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaofang Wang, Xiaoqing Ellen Tan, Xide Xia, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aayushi Srivastava, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Amos Teo, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Dong, Annie Franco, Anuj Goyal, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Ce Liu, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Cynthia Gao, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Eric-Tuan Le, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Filippos Kokkinos, Firat Ozgenel, Francesco Caggioni, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hakan Inan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Hongyuan Zhan, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Ilias Leontiadis, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Janice Lam, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kiran Jagadeesh, Kun Huang, Kunal Chawla, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Miao Liu, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikhil Mehta, Nikolay Pavlovich Laptev, Ning Dong, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Rangaprabhu Parthasarathy, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Russ Howes, Ruty Rinott, Sachin Mehta, Sachin Siby, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Mahajan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shishir Patil, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Summer Deng, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Koehler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaojian Wu, Xiaolan Wang, Xilun Wu, Xinbo Gao, Yaniv Kleinman, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yu Zhao, Yuchen Hao, Yundi Qian, Yunlu Li, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, Zhiwei Zhao, and Zhiyu Ma. 2024. The Llama 3 herd of models. *Preprint*, arxiv:2407.21783 [cs].

John T. Hale. 2014. *Automaton theories of human sentence comprehension*. CSLI Publications, Center for the Study of Language and Information.

Mantaro Hashimoto. 1969. Observations on the passive construction. *Unicorn*, 5(1):59–71.

Mantaro Hashimoto. 1987. The historical and geographical development of Chinese passive constructions. *Zhongguo Yuwen*, 196:36–49.

Cheng-Teh James Huang. 1982. *Logical relations in*

Chinese and the theory of grammar. Ph.D. thesis, Massachusetts Institute of Technology.

Alexander G. Huth, Wendy A. de Heer, Thomas L. Griffiths, Frédéric E. Theunissen, and Jack L. Gallant. 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600):453–458.

Paul Ibbotson. 2022. *Language acquisition: the basics*. Routledge.

Ray Jackendoff. 1977. *X-bar-syntax: a study of phrase structure*. MIT press.

Evan Kidd, Elena V. M. Lieven, and Michael Tomasello. 2010. Lexical frequency and exemplar-based learning effects in language acquisition: evidence from sentential complements. *Language Sciences*, 32(1):132–142.

Jixing Li, Shohini Bhattasali, Shulin Zhang, Berta Franzluebbers, Wen-Ming Luh, R. Nathan Spreng, Jonathan R. Brennan, Yiming Yang, Christophe Pallier, and John Hale. 2022. Le petit prince multilingual naturalistic fmri corpus. *Scientific Data*, 9(1):530.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60. Association for Computational Linguistics.

Eric Maris and Robert Oostenveld. 2007. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190.

Matthew J. Nelson, Imen El Karoui, Kristof Giber, Xiaofang Yang, Laurent Cohen, Hilda Koopman, Sydney S. Cash, Lionel Naccache, John T. Hale, Christophe Pallier, and Stanislas Dehaene. 2017. Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, 114(18):E3669–E3678.

Martin Schrimpf, Idan Asher Blank, Greta Tuckute, Carina Kauf, Eghbal A. Hosseini, Nancy Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko. 2021. The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, 118(45):e2105646118.

Einat Shetreet and Naama Friedmann. 2014. The processing of different syntactic structures: fMRI investigation of the linguistic distinction between *wh*-movement and verb movement. *Journal of Neurolinguistics*, 27(1):1–17.

Jon Sprouse, Matt Wagers, and Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language*, 88(1):82–123.

Mariya Toneva, Tom M. Mitchell, and Leila Wehbe. 2022. Combining computational controls with natural text reveals aspects of meaning composition. *Nature Computational Science*, 2(11):745–757.

Peter Chin-Tang Wang. 1970. *A Transformational Approach to Chinese Ba and Bei*. Ph.D. thesis, The University of Texas at Austin.

Appendix

A X-bar and trace annotations in English and Chinese

Table 2 lists the Chinese-specific functional categories. Table 3 reports the counts of different trace types derived from movement in English and Chinese. Figure 6 illustrates representative examples of movement in English and Chinese using X-bar tree structures.

B fMRI data acquisition and preprocessing

English and Chinese MRI data were acquired with a 3 T GE Discovery MR750 scanner using a 32-channel head coil. Structural scans were acquired using a T1-weighted magnetization-prepared rapid gradient-echo sequence. Functional scans were obtained using a multi-echo echo-planar imaging (EPI) sequence (TR = 2000 ms; TEs = $[12.8, 27.5, 43]$ ms; flip angle = 77°; matrix size = 72×72; FoV = 240.0×240.0 mm; in-plane acceleration factor = 2; 33 axial slices; voxel size = $3.75 \times 3.75 \times 3.8$ mm). Each scanning run began with a trigger followed by an 8 s delay before stimulus onset. Anatomical and functional MRI data were preprocessed with fMRIPrep (v25.0.0; Esteban et al., 2020), with final resampling to the fsaverage5 surface performed in a single interpolation step using `mri_vol2surf`.

C Banded ridge methods

In the banded ridge regression, fMRI responses y were predicted as

$$\hat{y} = \sum_i K_i w_i \tag{1}$$

where $K_i = X_i X_i^\top$ and w_i are the corresponding kernel weights. Model fitting minimized the objective

$$\|y - \sum_i K_i w_i\|^2 + \sum_i \alpha_i \, w_i^\top K_i w_i \tag{2}$$

with independent ridge penalties α_i for each kernel. We used the `precomputed` kernel option and performed random search over $\alpha_i \in [10^0, 10^{20}]$, optimizing log-weights $\delta_i = -\log \alpha_i$ via cross-validation. Data were split into 90% training and 10% testing sets in temporal order. Per-kernel predictions $\hat{y}_i$ were obtained using `predict(split=True)`, and Pearson correlations between predicted and observed responses were computed as model-specific performance scores.

D Additional GLM results

Statistics for significant brain clusters associated with syntactic node counts and traces from the GLM analysis are reported in Table 4. Clusters and corresponding statistics for control variables are shown in Figure 7 and Table 5.

E Banded ridge results

Table 6 reports cluster statistics for deep-, surface-, and combined-order word embeddings in English and Chinese from the banded ridge regression analysis.

F Computational resources

All experiments were conducted on a high-performance computing (HPC) cluster with nodes equipped with dual AMD EPYC 7742 processors (64 cores per socket; 128 physical cores per node) and 512 GB of RAM. For each participant, the GLM analysis required approximately 0.5 CPU-hours, and the banded ridge regression required approximately 5 CPU-hours.

Phrase	Head / Marker	Core function and description
ModP	neng, hui, yinggai	Encodes deontic, epistemic, or dynamic modality; headed by modal markers and structurally dominates vP.
PolarP	shi bu shi, neng fou	Encodes affirmative, negative, or interrogative polarity, typically realized through polarity markers.
ShiP	shi	Copular phrase connecting subject and predicate, realizing equivalence, classification, or focus-related emphasis.
QP	san ge, dou, you de	Quantifies nominal or verbal predicates; includes both nominal and adverbial quantifier types.
LianP	lian	Marks extreme focus (e.g., *lian . . . dou/ye*); the focused element occupies the specifier position.
DeP	de	Functions as an attributive marker or nominalizer of predicative phrases via the structural particle *de*.
BaP	ba	Disposal construction moving the patient to the specifier position; dominates VP and co-occurs with resultative or perfective markers.
ClP	ben, ci, tian	Classifier phrase forming the core of number–classifier–noun structures; includes nominal (dominating NP) and verbal (dominating VP) types.
YouP	you	Expresses possession, existence, or attributive relations via the verb *you*.
ReP	result elements	Encodes resultative meaning by dominating result-denoting elements associated with an action.
SitP	situation markers	Encodes durative situational states of events; e.g., *zhao* in *Ta shui zhao le* indicates continuation of the sleep state.
AspP	le, zhe, guo	Encodes aspectual properties of events; aspect markers indicate completion, durativity, or experiential aspect and dominate VP.

Table 2: Chinese-specific functional phrase categories.

Trace type	English	Chinese	Trace type	English	Chinese
t-DP	2330	1982	*t*-ClP	/	2
t-TP	8	1	*t*-C	245	/
t-T	1370	/	*t*-AspP	/	2
t-DeP	/	101	*t*-CP	128	188
t-V	216	428	*t*-Shi	/	1
t-GP	/	22	*t*-Op	117	/
t-v	219	269	*t*-AP	2	/
t-NumP	/	7	*t*-AdvP	134	1
t-PRO	328	214	*t*-VP	1	/
t-ConjP	1	6	*t*-PP	23	27
t-pro	46	183	*t*-NegP	1	1

Table 3: Counts of different types of traces in English and Chinese sentences.

Regressor	English			Chinese		
	N vertices	p	Cohen's d	N vertices	p	Cohen's d
Trace	731	0	2.9374	316	0.001	4.4334
X-bar bottom-up	111	0.0438	3.1838	300	0.0009	2.9952
X-bar top-down	419	0	2.6363	144	0.0679	4.4474
CFG bottom-up	474	0.0004	2.7472	207	0.0168	3.5706
CFG top-down	316	0.0220	2.2983	189	0.0469	3.5454
X-bar $>$ CFG						
Bottom-up	457	0.0415	2.1765	300	0.0009	2.9952
Top-down	511	0	3.2521	67	0.0338	2.8111

Table 4: Cluster statistics for syntactic node-count and trace effects from the GLM analysis in English and Chinese.

Control variable	English			Chinese		
	N vertices	p	Cohen's d	N vertices	p	Cohen's d
Pitch	465	0.0001	3.0234	561	0	1.9540
Intensity	373	0	2.6984	106	0.006	3.1360
Word rate	593	0	3.2627	579	0	2.1923
Frequency	690	0	-3.2925	231	0.0754	-1.9742

Table 5: Cluster statistics for control variables in English and Chinese from GLM analysis.

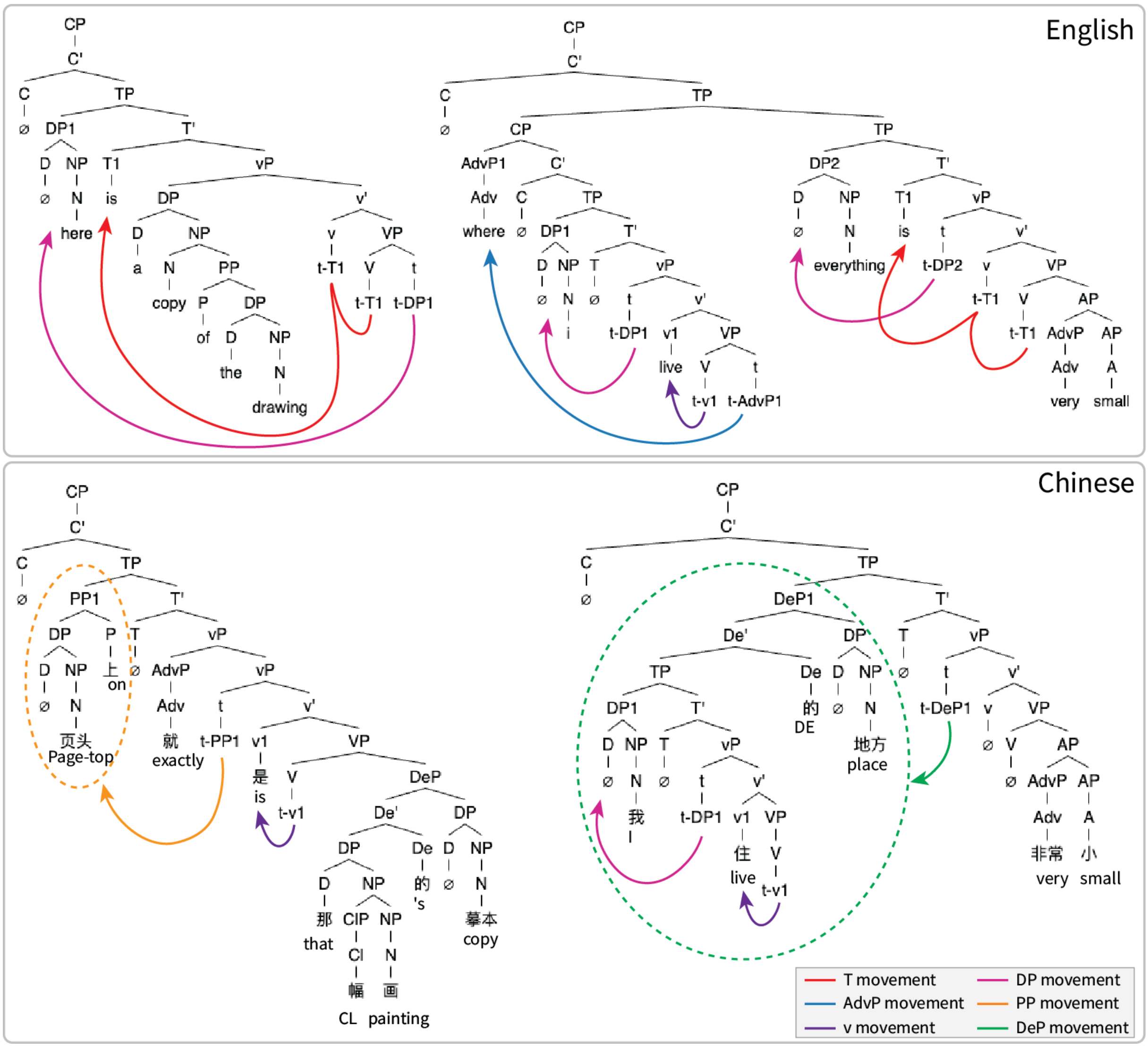

Figure 6: Representative examples of movement in English and Chinese using X-bar tree structures.

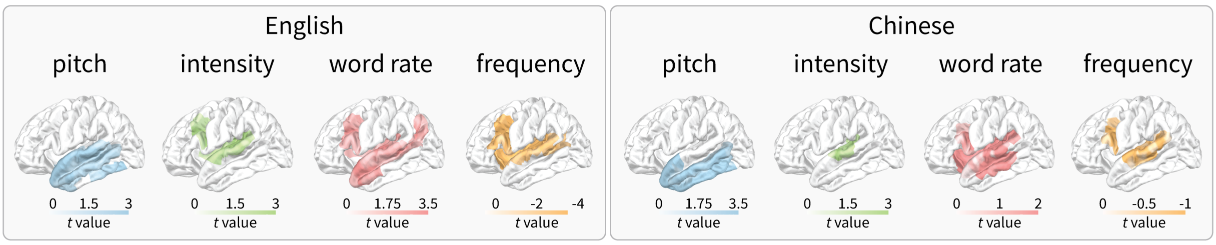

Figure 7: Significant clusters associated with control variables in English and Chinese from the GLM analysis.

Embedding group	English			Chinese		
	N vertices	***p***	**Cohen's *d***	**N vertices**	***p***	**Cohen's *d***
Deep	41	0.0095	4.0586	/	/	/
Surface	68	0.0434	2.5661	12	0.0300	2.3576
Combined	120	0.0498	1.9537	59	0.042	1.9463
Contrast						
Surface > Deep	64	0.0498	1.8627	44	0.0013	2.3757
Combined > Surface	17	0.0138	3.5557	/	/	/
Combined > Deep	58	0.0132	1.6109	12	0.0463	5.3000

Table 6: Cluster statistics for deep-, surface-, and combined-order word embeddings in English and Chinese from banded ridge regression.