Navigating Brain Language Representations: A Comparative Analysis of Neural Language Models and Psychologically Plausible Models

Yunhao Zhang[1,2], **Shaonan Wang**[1,2], **Xinyi Dong**[3], **Jiajun Yu**[4], **Chengqing Zong**[1,2]
[1]State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, CAS, Beijing, China
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[3]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University
[4]College of Information and Electrical Engineering, China Agricultural University
zhangyunhao2021@ia.ac.cn; {shaonan.wang, cqzong}@nlpr.ia.ac.cn;
202121061097@mail.bnu.edu.cn; 2017307070311@cau.edu.cn

Abstract

Neural language models, particularly large-scale ones, have been consistently proven to be most effective in predicting brain neural activity across a range of studies. However, previous research overlooked the comparison of these models with psychologically plausible ones. Moreover, evaluations were reliant on limited, single-modality, and English cognitive datasets. To address these questions, we conducted an analysis comparing encoding performance of various neural language models and psychologically plausible models. Our study utilized extensive multi-modal cognitive datasets, examining bilingual word and discourse levels. Surprisingly, our findings revealed that psychologically plausible models outperformed neural language models across diverse contexts, encompassing different modalities such as fMRI and eye-tracking, and spanning languages from English to Chinese. Among psychologically plausible models, the one incorporating embodied information emerged as particularly exceptional. This model demonstrated superior performance at both word and discourse levels, exhibiting robust prediction of brain activation across numerous regions in both English and Chinese.

Keywords: Neural Language Models, Psychologically Plausible Models, Neural Encoding

Introduction

Neural language models, particularly large ones, exhibit remarkable effectiveness in diverse downstream tasks and exceptional language understanding abilities. This success has spurred the development of studies utilizing neural language models to investigate how the human brain processes languages. Previous research utilizes representations generated by distinct models to predict brain activation (Mitchell et al., 2008; Huth, De Heer, Griffiths, Theunissen, & Gallant, 2016), assuming that the closer the representation aligns with the brain's semantic information, the better it captures brain activation.

Recent research has compellingly demonstrated that neural language models excel in predicting brain neural activities (Schrimpf et al., 2021). However, existing studies have overlooked the comparison with simple yet psychologically plausible models, which possess mathematically transparent computational mechanisms and do not necessitate extensive hyperparameter tuning processes. Furthermore, these studies often rely on a single modality of brain activity data from small cognitive datasets to evaluate models. These datasets typically involve stimuli from a single paradigm and a single language unit (e.g., word-level). Moreover, the focus of these investigations has primarily been on Germanic languages, particularly English, while Tibetan languages like Chinese remain largely unexplored.

Therefore, the effectiveness of neural language models in predicting brain activations and their superiority over other models, as well as their ability to capture the nuances of human language processing, remains uncertain. Comprehensive studies are essential to understand the intricate relationship between artificial intelligence and the brain's cognitive functions.

To bridge this gap, we conduct a comparative analysis, evaluating the encoding performance of diverse neural language models (NLMs) and psychologically plausible models (PPMs) on eight multi-modal cognitive datasets in Chinese and English, encompassing both word and discourse levels. To investigate spatial difference of model encoding performance, we perform a fine-grained analysis at both the region of interest (ROI) and voxelwise levels using fMRI data. Our findings reveal that: (i) Simple PPMs outperform NLMs across various contexts, encompassing diverse modalities like fMRI and eye-tracking and spanning languages from English to Chinese. (ii) PPMs integrating embodied and network-topological information excel in word-level encoding, with embodied models outperforming others at the discourse level. PPMs incorporating local-statistical information are particularly adept at fitting eye-tracking patterns. (iii) The shallow layers of NLMs excel in word-level brain activation, whereas middle layers are better at capturing discourse-level activation. (iv) The brain cortex encoding map reveals unique correlations between different models and various brain regions, implying distinct and exclusive information encoding within different models.

Related Work

Evaluating Neural Language Models with Cognitive Data

Previous research has attempted to identify models that more effectively capture brain activation. Abnar, Ahmed, Mijnheer, and Zuidema (2017) evaluates seven NLMs, encompassing distributional, dependency-based, and one experiential model, in predicting neural responses to 60 nouns presented by Mitchell et al. (2008). Beinborn, Abnar, and Choenni (2019) evaluates the ability of ELMo to predict brain responses on four small fMRI datasets. These studies

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

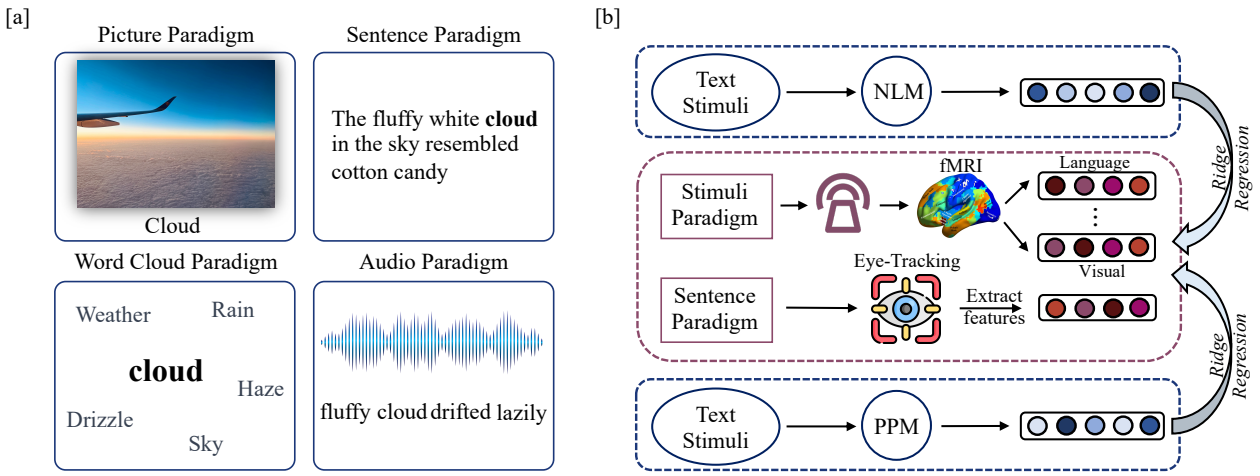

Figure 1: [a] Stimuli examples from four paradigms: picture, sentence, and word cloud paradigms are employed in word-level fMRI, while the audio paradigm is used in discourse-level fMRI. Sentence paradigm is utilized in eye-tracking data. [b] Neural encoding method using representations generated by neural language models and psychologically plausible models.

demonstrate NLMs' effective prediction of brain activation. Hollenstein, de la Torre, Langer, and Zhang (2019) evaluates six NLMs consisting of four context-independent models and two context-aware models on relatively more cognitive datasets. Moreover, Schrimpf et al. (2021) compares multiple state-of-art models on relatively small sentence-level datasets, from no more than 9 participants. These studies conclude that context-aware models perform best in encoding English cognitive data.

However, recent work has demonstrated that learning and evaluation goals commonly utilized with NLMs, like masked or next-word prediction, are inconsistent with the brain's language understanding mechanisms (Pasquiou, Lakretz, Hale, Thirion, & Pallier, 2022; Antonello & Huth, 2022). Therefore, NLMs may not be the optimal model for capturing brain activity. Moreover, cognitive datasets used in these studies are either too small or only contain stimuli from single paradigm and single language unit. Furthermore, these studies have predominantly centered around Germanic languages, particularly English, while Tibetan languages, such as Chinese, remain largely unexplored. Investigating the relationship between model representations and brain activation in diverse languages can enrich our comprehension of brain language representations (Wang et al., 2024).

Psychologically Plausible Models

In psycholinguistics, three distinct types of computational principles are considered as candidates for the semantic representational algorithm in the human brain.

The first type is local-statistical system, such as simple co-occurrence, whose variations can be detected by humans during language acquisition (Saffran, Senghas, & Trueswell, 2001; Conway & Christiansen, 2005). Several studies have found that co-occurrence information in language corpora predicts various human semantic phenomena (Roelke et al., 2018; Hofmann et al., 2018; Frank & Willems, 2017), indicating that co-occurrence information likely plays a pivotal role in shaping human semantic organization.

The second type is global-network-topological system. In network sciences, language can be conceptualized as a complex network, with words as nodes and their correlations as edges, displaying rich topological properties. Numerous studies have shown that humans implicitly infer these topological properties during various structural learning tasks, including motor sequence learning (Lynn, Kahn, Nyema, & Bassett, 2020), object relation learning (Garvert, Dolan, & Behrens, 2017), visual event segmentation (Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013), and picture naming (Fu et al., 2023).

The third type is embodied-based system. It is motivated by grounded semantics, where experiential information from diverse modality-specific systems, such as visual, motor, and social, is re-encoded into semantic representations stored in memory. Several studies have found that embodied-based system partly reflects conceptual knowledge in human brain (Damasio, 1989; Glenberg, 1997; Binder & Desai, 2011; Fernandino, Tong, Conant, Humphries, & Binder, 2022).

The aforementioned studies primarily investigate the relationship between PPMs and brain's neural representations at word level. It remains unclear to what extent PPMs encode the brain's neural representations of sentences and discourse, along with behavioral signals like eye-tracking. Furthermore, these studies do not incorporate NLMs for a comprehensive comparison and analysis.

Cognitive Datasets

We introduce the sources of English and Chinese cognitive datasets (See Table 1 for details.). For word-level fMRI data, we preprocess it using fMRIPrep (Esteban et al., 2019) and conduct first-level analysis to obtain t-value images rep-

Language	Modality	Source	Paradigm	Subject	Unit	Tokens
English	Word fMRI	Pereira et al. (2018)	Picture	15	Word	180
	Word fMRI	Pereira et al. (2018)	Text	15	Word	180
	Word fMRI	Pereira et al. (2018)	Word Cloud	15	Word	180
	Discourse fMRI	Y. Zhang, Han, Worth, and Liu (2020)	Audio	19	Discourse	47,356
	Eye-tracking	Hollenstein et al. (2018)	Text	12	Sentence	36,767
Chinese	Word fMRI	Wang, Zhang, Zhang, Sun, et al. (2022)	Picture	11	Word	672
	Discourse fMRI	Wang, Zhang, Zhang, and Zong (2022)	Audio	12	Discourse	52,269
	Eye-tracking	G. Zhang et al. (2022)	Text	1718	Sentence	170,331

Table 1: Details of the cognitive datasets used in our experiments.

resenting neural activation for each word. Discourse-level fMRI data is preprocessed using the Human Connectome Project (HCP) pipeline. To align word representations with fMRI data, the representations are convolved with the hemodynamic response function (HRF) and down-sampled to the discourse-fMRI sampling rate. In eye-tracking data, four key word-level features are extracted, encompassing the entire reading process. These features, categorized into total reading time (TRT) and gaze duration (GD), number of fixations (nFixations), and first fixation duration (FFD), capture various aspects of processing.

Method

In this section, we demonstrate how the representations produced by neural language models and psychologically plausible models are used to encode cognitive data.

Neural Language Models

We adopt six typical neural language models that can be divided into two groups: one is context-independent models including GloVe (Pennington, Socher, & Manning, 2014) and Word2Vec (Mikolov, Chen, Corrado, & Dean, 2013), where GloVe is a count-based method that performs a dimensionality reduction on the co-occurrence matrix, and Word2Vec employs a shallow neural network to map words from a large corpus into continuous vector spaces. The other is context-aware models including GPT2_{ori}, GPT2_{med}, $\text{BERT}_{\text{base}}$ and BERT_{lar}, where GPT2 is an auto-regressive language model, trained to predict the next token based on preceding text, and BERT is an auto-encoder language model, trained bidirectionally to predict masked tokens.

To obtain word embeddings, we train GloVe and Word2Vec on large-scale corpora, the Xinhua News corpus (19.7 GB)[1] for Chinese and the Wikipedia corpus (13 GB)[2] for English, using identical model parameters, including the Skip-Gram architecture, a negative number of 15 for Word2Vec, a window width of 2, and embedding dimensions of 300. GPT2_{ori} and $\text{BERT}_{\text{base}}$ have 12 hidden layers, GPT2_{med} and $\text{BERT}_{\text{large}}$ have 24 hidden layers, and we take word representations from each hidden layer. We derive each word's representation by averaging sub-word embeddings for discourse stimuli and adopt Chersoni, Santus, Huang, Lenci, et al. (2021) and Y. Zhang, Li, Zhang, Dong, and Wang (2023)'s methods to extract word representations for word stimuli.

Psychologically Plausible Models

We developed three psychologically plausible models, each based on a unique computational principle for semantic representation in the human brain.

Local-Statistical Model (LSM) Following previous studies (Fu et al., 2023; Frank & Willems, 2017), we construct the local-statistical model by calculating a co-occurrence matrix weighted by Positive Pointwise Mutual Information and reducing it with Singular Value Decomposition. Finally, we obtain a 300-dimensional representation for each word.

Network-Topological Model (NTM) Inspired by previous studies (Newman-Griffis & Fosler-Lussier, 2017), we first calculate the one-order similarity coefficients (e.g., cosine similarity) among words. Then, we construct a graph, where each node is a word and each edge is cosine similarity coefficients. Finally, we use random walks and dimensionality reduction techniques to generate network-topological word representations.

Embodied-Based Model (EBM) Word representations with six semantic dimensions were obtained from a prior study (Wang et al., 2023), which represent various aspects of word meanings, such as vision, motor, socialness, emotion, time, and space, in alignment with neural processing systems (Binder et al., 2016). We obtain a 6-dimensional representation encompassing both sensory-motor and abstract semantic information for each word.

Encoding Brain Activation

As shown in Figure 1[b], to map representations yielded by different models (NLMs and PPMs) to brain activations, for each cognitive dataset, we train encoding models to predict fMRI or eye-tracking signals from text stimuli for each subject. Specifically, we follow K-fold (K=10) cross-validation. All data samples from K-1 folds were utilized for training, while the model is tested on the remaining fold. Following previous works (Oota et al., 2022), we employ sklearn's ridge regression, 10-fold cross-validation, mean squared error loss function. Moreover, we perform a group-level paired t-test with false discovery rate (FDR) correction to assess the significance of comparison results.

[1] http://www.xinhuanet.com/whxw.htm

[2] https://dumps.wikimedia.org/enwiki/latest

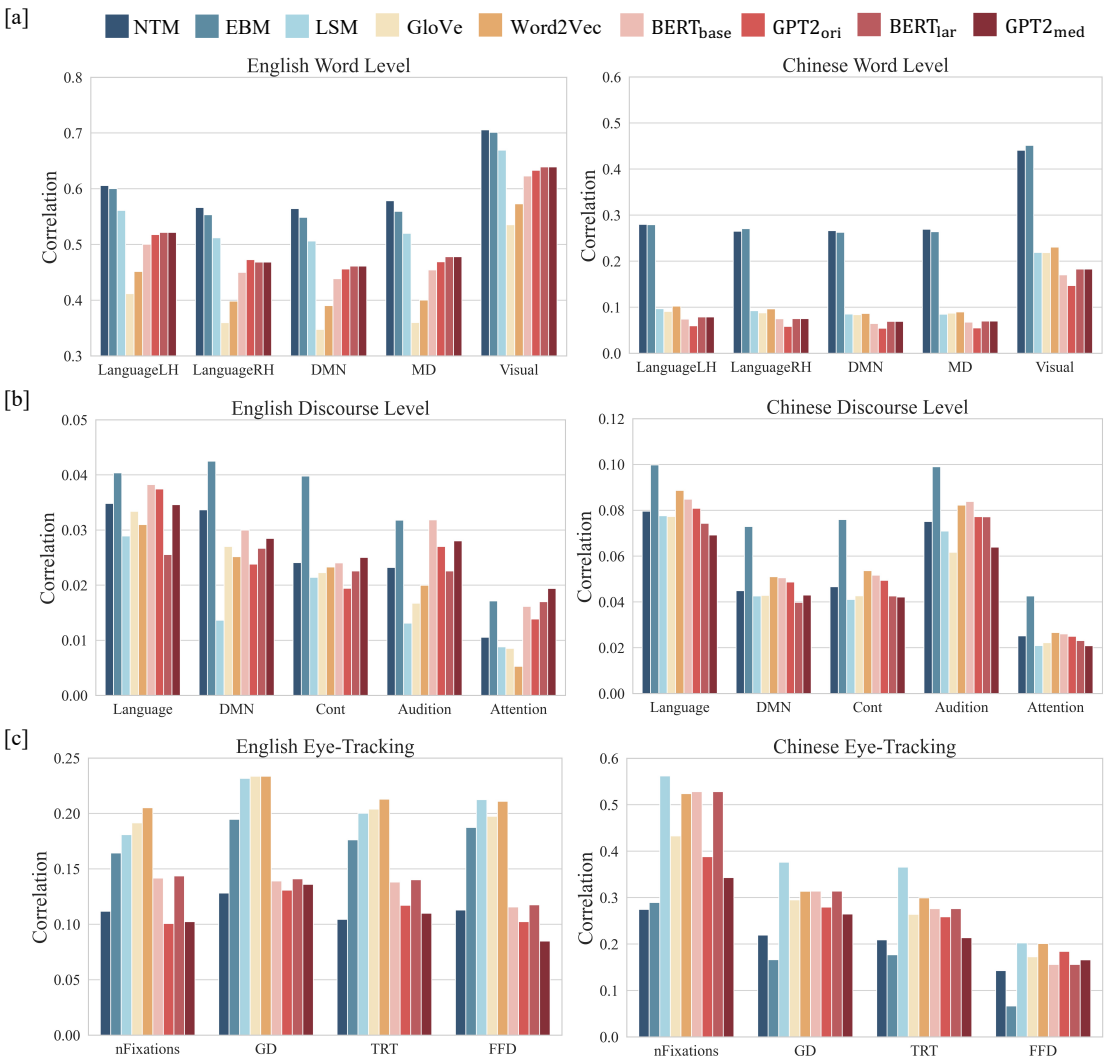

Figure 2: Pearson correlation coefficients between predicted and true values were computed for English and Chinese in word-level fMRI, discourse-level fMRI, and eye-tracking data using both NLMs and PPMs. Results are averaged across all subjects. To facilitate comparison, we average results from the picture paradigm, sentence paradigm, and word cloud paradigm to obtain the English word-level results. As for context-aware models, we select the layer with the best performance.

Result and Discussion

Figure 2 illustrates the mean encoding performance of NLMs and PPMs across subjects for selected brain networks in Chinese and English, covering word-level fMRI, discourse-level fMRI and eye-tracking data. Figure 3 displays the mean performance of each layer within context-aware models across whole brain for both Chinese and English in word-level and discourse-level fMRI. Figure 4 shows distribution of top-performing model for each voxel across whole brain.

Comparison between Neural Language Models and Psychologically Plausible Models

Figure 2 illustrates that at the word level, PPMs significantly outperform NLMs across various brain networks and languages like English and Chinese ($p < 0.05$, FDR corrected). At the discourse level, PPMs show superior average performance across most brain networks ($p < 0.05$, FDR corrected). However, in brain networks with less emphasis on language processing (Audition and Attention), NLMs exhibit comparable performance to PPMs in English. In eye-tracking, PPMs have comparable performance to NLMs on average ($p > 0.05$, FDR corrected). Moreover, Figure 3 indicates that at the word level, PPMs outperform each layer of context-aware models ($p < 0.05$, FDR corrected). In the discourse level, PPMs on average outperform NLMs across the entire brain. In summary, PPMs prove to be more effective in predicting brain activation than NLMs.

In word level, the better performance in PPMs supports popular research views that network-topological and embodied properties explain certain aspects of the brain's concep-

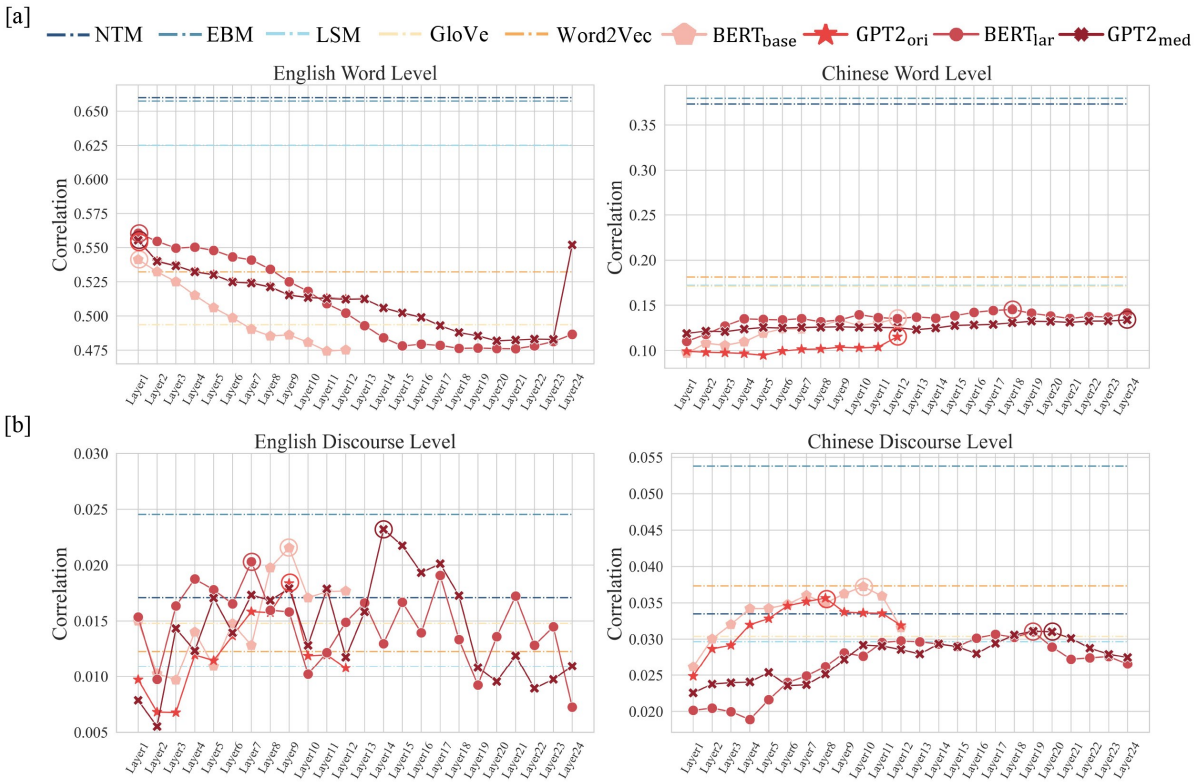

Figure 3: Mean Pearson correlation coefficients were calculated for each layer of context-aware models in English and Chinese word-level and discourse-level fMRI data across the entire brain. A circle marks the layer with the best encoding performance for each model.

tual representation mechanism, respectively (Fu et al., 2023; Fernandino et al., 2022). The lower performance in NLMs may be attributed to the learning and evaluation goals frequently used with NLMs, such as masked or next-word prediction, which are inconsistent with the brain's language comprehension mechanisms. Consequently, NLM-generated representations lack similarity to human brain semantic information. Furthermore, the performance of context-aware models is relatively close to PPMs at the discourse level compared to the word level. It's probably because context-aware models can obtain the specific meaning of a word within its context. Therefore, they encode brain activation for complex language unit containing context, such as discourse, more effectively.

Comparison within Psychologically Plausible Models

Word Level Both EBM and NTM obtain best performance in predicting word-level brain activation in English and Chinese. And in Figure 2[a], LSM performs similarly to NTM and EBM in the English paradigm but lags behind in the Chinese paradigm. Compared to English, Chinese has more intricate grammar and text structure (Tang, 2021). Therefore, the simple local-statistical model like LSM struggles to capture word-level brain activation from Chinese text corpus.

Discourse Level Figure 2[b] illustrates EBM's superior encoding performance in various brain networks. This can be attributed to the situational nature of discourse-level tasks (Speelman & Kirsner, 1990), which require higher-order cognitive functions like imaginative scenario construction and benefit from more sensory-motor information involvement. Moreover, LSM obtains the lowest performance in PPMs, indicating that local-statistical representations can not effectively capture brain activation during comprehending content at discourse level.

Eye-Tracking As shown in Figure 2[c], LSM outperforms other PPMs on nearly all eye-tracking features, in line with psychological research indicating that local statistical patterns, such as co-occurrence, can predict reading behavior, which is measured using eye-tracking data (McDonald & Shillcock, 2003). Moreover, NTM and EBM achieve relatively low encoding performance, suggesting that effectively modeling brain activation may not achieve the same effect for behavioral signals.

Comparison within Neural Language Models

Word Level As shown in Figure 2[a], we observe that context-aware models have better encoding performance than that of context-independent models in the English paradigm, while context-independent models excel in context-aware models in the Chinese paradigm. In psycholinguistics, the holistic view suggests that due to the prevalence of homographic and homophonic morphemes in Chinese, holistic processing of compound words is more efficient (Packard, 1999).

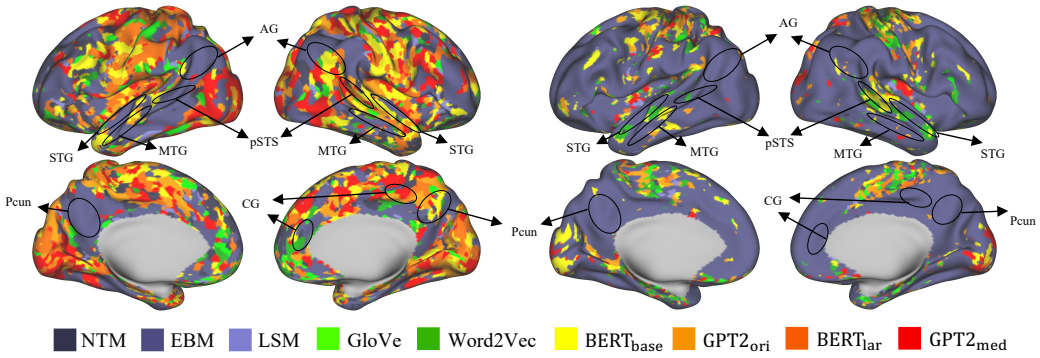

Figure 4: Distribution of top-performing model for each voxel across the entire brain in English (left) and Chinese (right) discourse-level fMRI. ROIs related to semantic processing are marked.

And neuroimaging studies have highlighted the brain's reliance on holistic word-level processing during Chinese semantic learning (Tsang & Zou, 2022). Therefore, BERT and GPT2 models in Chinese acquire semantic information through character-level training, diverging from human brain processing. Yet Word2Vec in Chinese learns word-level representations from the large-scale corpus, which have better encoding performance than context-aware models.

As shown in Figure 3[a], in the Chinese paradigm, context-aware models show consistent encoding performance across various layer depths, whereas in the English paradigm, they exhibit significant variation across various layer depths. This result suggests that, in English context-aware models, more semantic information akin to the human brain is present in shallow layers, whereas in Chinese context-aware models, it is evenly distributed across layers. Furthermore, GPT2$_{med}$ and BERT$_{large}$ has better performance than GPT2$_{origin}$ and BERT$_{base}$. This result suggests that context-aware models with more parameters capture additional word-level semantic information akin to the human brain.

Discourse Level As observed in Figure 2[b], context-aware models outperform context-independent models in English, while the context-independent model (Word2Vec) excels context-aware models in Chinese, in line with word-level findings.

As shown in Fig 3[b], context-aware models show a rise followed by a decline in performance as the layer rises, which is similar on English and Chinese datasets. However, the shallow layers of context-aware models have higher performance than the deep layers on English word-level datasets. These language-unit differences between layers suggest that, in context-aware models, shallow layers encode simple linguistic units, while middle layers capture richer semantic information related to more complex linguistic units (Y. Zhang, Zhang, Li, Wang, & Zong, 2024).

Eye-Tracking It can be noticed from Figure 2[c] that in English, context-independent models outperform context-aware models, contrary to fMRI results. It indicates that effectively modeling brain activation may not yield the same impact on behavioral signals.

Moreover, NLMs accurately predict TRT and FFD in English, while they predict TRT better than FFD in Chinese, which may indicate that Chinese models capture the information of late-stage semantic integration better than early processing stages of lexical access.

Encoding Performance on Each Voxel

To thoroughly explore encoding performance across various models, we label each voxel using the model exhibiting the best performance for that voxel. As illustrated in Figure 4, EBM consistently achieves superior encoding performance in nearly all ROIs related to semantic processing, highlighting its ability to capture brain activation during language comprehension. Furthermore, we observe that distinct models capture activation from different brain regions. For example, GPT2$_{med}$ captures activation in regions such as the lateral occipital cortex (LocG), insular gyrus (INS), and superior parietal lobule (SPL), while BERT$_{base}$ captures activation in regions including the angular gyrus (AG), middle frontal gyrus (MFG), posterior superior temporal sulcus (pSTS), middle temporal gyrus (MTG), and precentral gyrus (PrG). The cortical encoding map indicates unique correlations between specific models and various brain regions, suggesting distinct and exclusive information encoding within these models.

Conclusion

Analyzing multi-modal cognitive data across diverse contexts and languages, our study shows that psychologically plausible models outperform neural language models in brain encoding performance, challenging current claims of neural models' excellence in predicting brain activation during language processing. Moreover, our results replicate existing findings regarding the encoding performance of neural language models with English cognitive data, as well as discover the prominent contribution of embodied and network-topological information to brain activation prediction in both English and Chinese.

Our findings offer valuable insights for selecting computational models in diverse cognitive tasks, shedding light on language processing across languages and contexts, guiding future research.

Acknowledgements

We would like to thank the anonymous reviewers for their valuable comments. This research was supported by grants from the National Natural Science Foundation of China to S.W. (62036001) and S.W. (the STI2030-Major Project, grant number: 2021ZD0204105).

References

Abnar, S., Ahmed, R., Mijnheer, M., & Zuidema, W. (2017). Experiential, distributional and dependency-based word embeddings have complementary roles in decoding brain activity. *arXiv preprint arXiv:1711.09285*.

Antonello, R., & Huth, A. (2022). Predictive coding or just feature discovery? an alternative account of why language models fit brain data. *Neurobiology of Language*, 1–16.

Beinborn, L., Abnar, S., & Choenni, R. (2019). Robust evaluation of language–brain encoding experiments. In *International conference on computational linguistics and intelligent text processing* (pp. 44–61).

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive neuropsychology*, *33*(3-4), 130–174.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in cognitive sciences*, *15*(11), 527–536.

Chersoni, E., Santus, E., Huang, C.-R., Lenci, A., et al. (2021). Decoding word embeddings with brain-based semantic features. *Computational Linguistics*, *47*(3), 663–698.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1-2), 25–62.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., ... others (2019). fmriprep: a robust preprocessing pipeline for functional mri. *Nature methods*, *16*(1), 111–116.

Fernandino, L., Tong, J.-Q., Conant, L. L., Humphries, C. J., & Binder, J. R. (2022). Decoding the information structure underlying the neural representation of concepts. *Proceedings of the National Academy of Sciences*, *119*(6), e2108091119.

Frank, S. L., & Willems, R. M. (2017). Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Language, Cognition and Neuroscience*, *32*(9), 1192–1203.

Fu, Z., Wang, X., Wang, X., Yang, H., Wang, J., Wei, T., ... Bi, Y. (2023). Different computational relations in language are captured by distinct brain systems. *Cerebral Cortex*, *33*(4), 997–1013.

Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *elife*, *6*, e17086.

Glenberg, A. M. (1997). What memory is for. *Behavioral and brain sciences*, *20*(1), 1–19.

Hofmann, M. J., Biemann, C., Westbury, C., Murusidze, M., Conrad, M., & Jacobs, A. M. (2018). Simple co-occurrence statistics reproducibly predict association ratings. *Cognitive science*, *42*(7), 2287–2312.

Hollenstein, N., de la Torre, A., Langer, N., & Zhang, C. (2019). Cognival: A framework for cognitive word embedding evaluation. *arXiv preprint arXiv:1909.09001*.

Hollenstein, N., Rotsztejn, J., Troendle, M., Pedroni, A., Zhang, C., & Langer, N. (2018). Zuco, a simultaneous eeg and eye-tracking resource for natural sentence reading. *Scientific data*, *5*(1), 1–13.

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458.

Lynn, C. W., Kahn, A. E., Nyema, N., & Bassett, D. S. (2020). Abstract representations of events arise from mental errors in learning and memory. *Nature communications*, *11*(1), 2313.

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological science*, *14*(6), 648–652.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *science*, *320*(5880), 1191–1195.

Newman-Griffis, D., & Fosler-Lussier, E. (2017). Second-order word embeddings from nearest neighbor topological features. *arXiv preprint arXiv:1705.08488*.

Oota, S. R., Arora, J., Agarwal, V., Marreddy, M., Gupta, M., & Surampudi, B. R. (2022). Neural language taskonomy: Which nlp tasks are the most predictive of fmri brain activity? *arXiv preprint arXiv:2205.01404*.

Packard, J. L. (1999). Lexical access in chinese speech comprehension and production. *Brain and Language*, *68*(1-2), 89–94.

Pasquiou, A., Lakretz, Y., Hale, J., Thirion, B., & Pallier, C. (2022). Neural language models are not born equal to fit brain data, but training helps. *arXiv preprint arXiv:2207.03380*.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1532–1543).

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., ... Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation.

Nature communications, *9*(1), 963.

Roelke, A., Franke, N., Biemann, C., Radach, R., Jacobs, A. M., & Hofmann, M. J. (2018). A novel co-occurrence-based approach to predict pure associative and semantic priming. *Psychonomic Bulletin & Review*, *25*, 1488–1493.

Saffran, J. R., Senghas, A., & Trueswell, J. C. (2001). The acquisition of language by children. *Proceedings of the National Academy of Sciences*, *98*(23), 12874–12875.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature neuroscience*, *16*(4), 486–492.

Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., ... Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, *118*(45), e2105646118.

Speelman, C. P., & Kirsner, K. (1990). The representation of text-based and situation-based information in discourse comprehension. *Journal of Memory and Language*, *29*(1), 119–132.

Tang, J. (2021). Differences between chinese and english thinking from the use of nouns in chinese and english. In *2nd international conference on language, art and cultural exchange (iclace 2021)* (pp. 20–25).

Tsang, Y.-K., & Zou, Y. (2022). An erp megastudy of chinese word recognition. *Psychophysiology*, *59*(11), e14111.

Wang, S., Sun, J., Zhang, Y., Lin, N., Moens, M.-F., & Zong, C. (2024). Computational models to study language processing in the human brain: A survey. *arXiv preprint arXiv:2403.13368*.

Wang, S., Zhang, X., Zhang, J., & Zong, C. (2022). A synchronized multimodal neuroimaging dataset for studying brain language processing. *Scientific Data*, *9*(1), 590.

Wang, S., Zhang, Y., Shi, W., Zhang, G., Zhang, J., Lin, N., & Zong, C. (2023). A large dataset of semantic ratings and its computational extension. *Scientific Data*, *10*(1), 106.

Wang, S., Zhang, Y., Zhang, X., Sun, J., Lin, N., Zhang, J., & Zong, C. (2022). An fmri dataset for concept representation with semantic feature annotations. *Scientific Data*, *9*(1), 721.

Zhang, G., Yao, P., Ma, G., Wang, J., Zhou, J., Huang, L., ... others (2022). The database of eye-movement measures on words in chinese reading. *Scientific Data*, *9*(1), 411.

Zhang, Y., Han, K., Worth, R., & Liu, Z. (2020). Connecting concepts in the brain by mapping cortical representations of semantic relations. *Nature communications*, *11*(1), 1877.

Zhang, Y., Li, C., Zhang, X., Dong, X., & Wang, S. (2023). A comprehensive neural and behavioral task taxonomy method for transfer learning in nlp. In *Findings of the association for computational linguistics: Ijcnlp-aacl 2023 (findings)* (pp. 233–241).

Zhang, Y., Zhang, X., Li, C., Wang, S., & Zong, C. (2024). Mulcogbench: A multi-modal cognitive benchmark dataset for evaluating chinese and english computational language models. *arXiv preprint arXiv:2403.01116*.