COMPARING CORTICAL ORGANIZATION AND SEMANTIC TUNING ACROSS BILINGUAL BRAINS

Anonymous authors
Paper under double-blind review

ABSTRACT

In bilingual individuals, the brain regions involved in processing semantics in native and non-native languages are broadly overlapping and represent largely similar semantic concepts, although fine-grained differences exist. However, how these broad similarities and fine-grained differences vary across individuals remains unclear. Here, we compared the spatial cortical organization and semantic tuning of semantic representations using functional magnetic resonance imaging (fMRI) data from six Chinese-English bilingual participants who read several hours of narratives in both languages. We fit voxelwise encoding models separately for each participant and language and applied hierarchical clustering to group-averaged estimated model weights to derive a common set of semantic clusters. We then compared the cortical organization and semantic tuning of these clusters across participants and languages. For all clusters, cortical overlap was higher across languages within a participant than across participants within a language. Semantic tuning similarity showed the opposite pattern, suggesting largely shared tuning across participants within the same language. Cross-participant semantic tuning similarity was lower for the communication and family clusters than for the other clusters, suggesting greater individual variability for social concepts. Together, these findings reveal a participant-specific spatial cortical organization that is preserved across languages and fine-grained language-specific patterns of semantic tuning largely shared across participants.

1 INTRODUCTION

Bilingual individuals can understand concepts in two languages. How the brain represents meaning across languages, and how this differs across individuals, is a central question in cognitive neuroscience. A large body of neuroimaging research in bilinguals suggests that representations of native and non-native languages are located in largely overlapping brain regions within individuals (Klein et al., 1995; Illes et al., 1999; Luke et al., 2002; Abutalebi et al., 2001; Perani & Abutalebi, 2005; Rüschemeyer et al., 2006; Yokoyama et al., 2006; Abutalebi & Green, 2007). However, some studies have found fine-grained differences in the precise cortical location (Dehaene et al., 1997; Kim et al., 1997; Xu et al., 2017). These brain regions also show broadly similar semantic tuning across languages (Honey et al., 2012; Buchweitz et al., 2012; Yang et al., 2017a;b; Dehghani et al., 2017) but fine-grained differences in tuning across languages have also been observed within individuals (Chen et al., 2026). However, how these similarities and differences compare with variations across individuals remains unclear.

To address this question, we compare semantic representations across participants and languages in bilingual brains, both in terms of their spatial cortical organization and semantic tuning. We analyzed functional magnetic resonance imaging (fMRI) data collected while six Chinese-English bilingual participants read several hours of narrative stories in both languages (Chen et al., 2026). We estimated fine-grained semantic representations separately for each participant and language using voxelwise encoding models (Naselaris et al., 2011). We identified cortical regions with similar semantic tuning using hierarchical clustering (Ward Jr, 1963; Meschke et al., 2023). We then systematically compared the cortical organization and semantic tuning of these clusters across participants and languages.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We find that semantic concepts are represented in overlapping brain regions across languages, but that these regions vary across participants. This is consistent with findings that cortical regions involved in semantic processing are language-agnostic (Malik-Moraleda et al., 2022), and that functional specialization and anatomy are idiosyncratic (Fedorenko & Kanwisher, 2009). We find the opposite pattern for semantic tuning, where representations are more similar across participants within one language than across languages within one participant. This suggests that the language-specific fine-grained differences that were previously observed (Chen et al., 2026) are consistent across participants. Moreover, while these patterns are consistent across clusters, we observe that semantic representations of social concepts appear to be more participant-specific (Visconti di Oleggio Castello et al., 2025). Overall, these findings reveal a distinction between participant-specific cortical organization shared across languages and language-specific patterns of semantic tuning shared across participants.

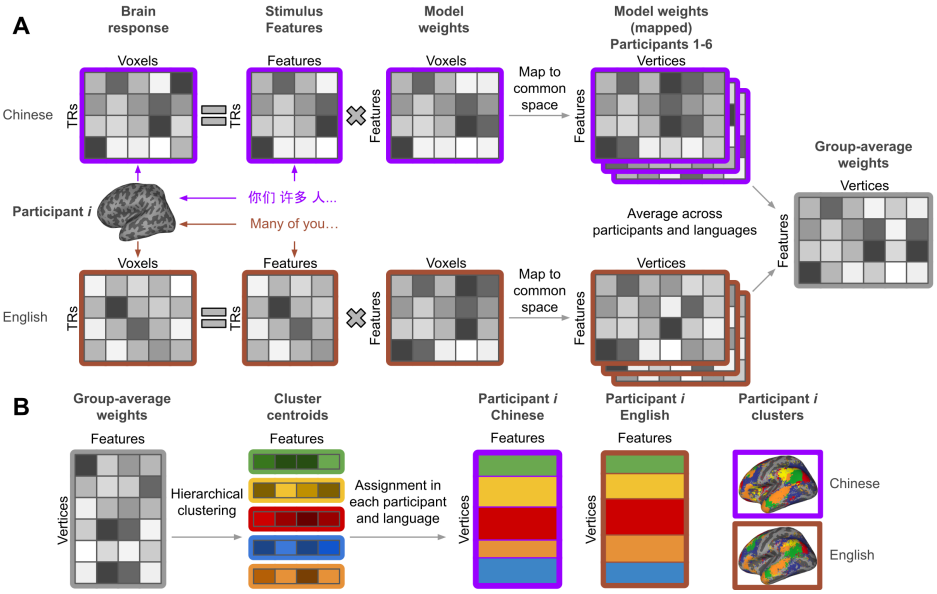

Figure 1: **Encoding framework and clustering procedure. A)** Voxelwise encoding models were trained in each participant and in Chinese and English separately. Brain responses were collected using fMRI while six bilingual participants read the same narratives in both languages. Features that represent the semantic content of the stimuli were extracted using multilingual BERT or fastText. Ridge regression was used to estimate model weights that describe voxelwise semantic tuning. These models were trained on a set of narratives and tested on a held-out narrative. The model weights of each participant in native brain space were projected to a common brain space (fsaverage). The projected weights were averaged across participants and languages to obtain group-average weights. **B)** Hierarchical clustering was performed on the group-average weights. Five clusters were retained and later interpreted. For each participant and language, vertices in the common space were assigned to the cluster whose centroid was closest.

2 Related work

Our work builds upon Chen et al. (2026), using the same bilingual naturalistic reading fMRI dataset and semantic clusters to examine how cortical organization and semantic tuning vary across participants and languages. They found that cluster assignments and semantic tuning are broadly consistent across languages, with language-specific fine-grained differences. However, their comparisons of semantic representations focused on differences across languages, rather than across participants. We extend this work by comparing both cortical organization and semantic tuning across participants. Botch & Finn (2024) compared representations of abstract and concrete concepts within and across participants during naturalistic listening in English. They found greater variability for abstract concepts, both within and across participants. While they examined within-participant stability of

semantic representations across stories, we examine within-participant similarities across languages, both in semantic tuning and cortical organization. Visconti di Oleggio Castello et al. (2025) used optimal transport to quantify individual differences in semantic representations during naturalistic listening in English. They identified participant-specific biases and found greater individual variability for social concepts than for physical concepts. Our work complements this approach by separately quantifying semantic tuning similarity and cortical overlap for corresponding semantic clusters, and by comparing these measures across both participants and languages.

3 METHODOLOGY

3.1 NATURALISTIC BRAIN IMAGING DATASET

Blood oxygen level-dependent (BOLD) responses were recorded from six Chinese-English bilingual participants using fMRI. All participants were native Mandarin Chinese speakers who learned English as a second language and were fluent daily users of both languages. Participants read the same stimuli in both languages: 11 narrative stories from *The Moth Radio Hour* originally in English and translated into Chinese by a professional speaker (Huth et al., 2016; Chen et al., 2026). Each narrative lasts between 10 to 15 minutes, resulting in over 2 hours of data in each language per participant. 10 narratives were used to estimate the encoding models and the 11th narrative was held out to evaluate the estimated models. Further details about the data collection and preprocessing can be found in Chen et al. (2026).

3.2 SEMANTIC FEATURES

We represented the semantic information of the stimulus words using a lexical (static) and a contextual embedding space. For the lexical embedding space, we used the fastText vectors aligned across languages (Joulin et al., 2018; Bojanowski et al., 2017). These vectors map each word onto a 300-dimensional semantic space. For the contextual embedding space, we used multilingual BERT (mBERT) (Devlin et al., 2019). mBERT is a 12-layer contextual language model with a hidden state size of 768. We obtained a contextual vector for each word by providing the corresponding stimulus sentence to mBERT and extracting the word representation from layer 9. For words that were split into several tokens, we averaged the representations of their tokens. We chose layer 9 because it produces the best aligned embeddings across languages (Chen et al., 2026) and because intermediate layers of contextual language models were shown to predict brain responses most accurately (Toneva & Wehbe, 2019; Caucheteux & King, 2022; Lamarre et al., 2022; Chen et al., 2024; Negi et al., 2025).

3.3 SENSORY-LEVEL FEATURES

We accounted for the effects of sensory information on brain responses by creating seven feature spaces that reflect low-level characteristics of the stimuli. These feature spaces represent the character count for Chinese or letter count for English, word count, single phonemes, diphones, triphones (Gong et al., 2023), spatiotemporal visual features (motion energy) (Adelson & Bergen, 1985; Watson & Ahumada Jr, 1985; Nishimoto & Gallant, 2011; Nunez-Elizalde et al., 2021) and intermediate features that capture orthographic similarities by measuring pixel overlap between words (Gong, 2024).

3.4 FEATURE SPACE PREPROCESSING

We aligned the feature spaces with the fMRI temporal resolution using a three-lobe Lanczos low-pass filter with a 0.25 Hz cut-off frequency. We accounted for the temporal delay between neural activity and the measured BOLD response by implementing a finite impulse response (FIR) model. The FIR model was created using time-shifted versions of each feature space at delays of 2, 4, 6, and 8 seconds in order to capture the hemodynamic response of individual voxels, following prior studies (Huth et al., 2016; Deniz et al., 2019; Nakai et al., 2021).

3.5 VOXELWISE ENCODING MODEL FITTING

We determined how features are represented in each voxel by estimating voxelwise encoding models (Wu et al., 2006; Naselaris et al., 2011). For each voxel, the models estimate regression weights that characterize the BOLD responses as a linear combination of features from a given feature space. We used banded ridge regression, a variant of standard ridge regression that allows different regularization parameters for each feature space and can therefore avoid biases caused by differences in feature distributions (Nunez-Elizalde et al., 2019). Banded ridge regression estimates the weights for each feature space by solving $\hat{B} = argmin_B \left\|Y - BF(X)\right\|_2^2 + \lambda \left\|CB\right\|_2^2$ where B is the mapping (dimension $v \times \sum_i^m p_i$), $F'(X)$ is the concatenated feature of m delayed feature spaces (each of dimension p_i), Y is the BOLD responses (dimension $v \times n$) and C the diagonal matrix of regularization parameters, with v the number of voxels and n the number of TRs. We used 5-fold cross-validation to optimize the C for each feature space and each voxel. We selected the hyperparameters that minimized the squared error by using a random search procedure with 1000 normalized hyperparameter candidates randomly sampled from a Dirichlet distribution and scaled by 21 log-spaced values ranging from 10^{-10} to 10^{10}. We fit the encoding models using the Himalaya Python package (La Tour et al., 2022).

3.6 STEPWISE REGRESSION PROCEDURE

We controlled for correlations between semantic and sensory-level feature spaces in the stimulus by using a stepwise regression procedure. We first estimated joint encoding models that predict BOLD responses from seven sensory-level features. We only used data from the train narratives to estimate the sensory-level models. We then predicted BOLD responses to the train and to the test narratives. We computed residual BOLD responses by subtracting the predicted BOLD responses $\hat{Y}_{lowlevel,train}$ and $\hat{Y}_{lowlevel,test}$ from the recorded BOLD responses Y_{train}, Y_{test}. For each voxel, we z-scored the residual BOLD responses $Y_{train} - \hat{Y}_{lowlevel,train}$ and $Y_{test} - \hat{Y}_{lowlevel,test}$ to unit variance centered around zero. We used the z-scored residual BOLD responses to estimate encoding models with the semantic feature spaces.

3.7 HIERARCHICAL CLUSTERING

To identify sets of voxels that have similar semantic tuning, we used hierarchical clustering (Ward Jr, 1963; Meschke et al., 2023). To enable comparisons across participants, we projected estimated semantic model weights of each participant and language onto a common brain space (fsaverage) (Fischl et al., 1999). We normalized projected model weights of each vertex to unit L2-norm. We averaged the normalized model weights across participants and languages, yielding one set of group-averaged weights in vertex space. We selected the vertices that were well-predicted in both languages (group-averaged $\sqrt{R^2_{zh}} > 0.05$ and $\sqrt{R^2_{en}} > 0.05$). We performed hierarchical clustering on the group-averaged weights of the well-predicted vertices. Following Chen et al. (2026), we used 5 clusters. This number of clusters was optimized through leave-one-out cross-validation in that study, evaluating how accurately group-level model weight clusters predicted BOLD responses for a held-out participant and language.

3.8 MEASURES OF SIMILARITY

To quantify cortical overlap, we compared three measures: Dice score, intersection over union (IoU) and match rate. All measures resulted in consistent findings. We therefore report the Dice score for consistency with previous studies Wilson et al. (2018); Khosla et al. (2021); Meschke et al. (2023); Thye et al. (2024). To quantify semantic tuning similarity, we used the Pearson correlation of mean semantic model weights (Deniz et al., 2019; Popham et al., 2021; Deniz et al., 2023; Lamarre et al., 2025).

4 RESULTS

To study how semantic representations differ across participants and languages, we compared semantic clusters in the brains of Chinese-English bilinguals. We identified these clusters from fMRI

data collected while six participants read several hours of narratives in both languages. We first estimated semantic model weights that reflect the tuning of each voxel, independently for each participant and language. We then applied hierarchical clustering to the group-averaged semantic model weights. We compared the resulting clusters separately in terms of spatial cortical organization and semantic tuning.

4.1 SEMANTIC CLUSTERS AND THEIR CORTICAL ORGANIZATION

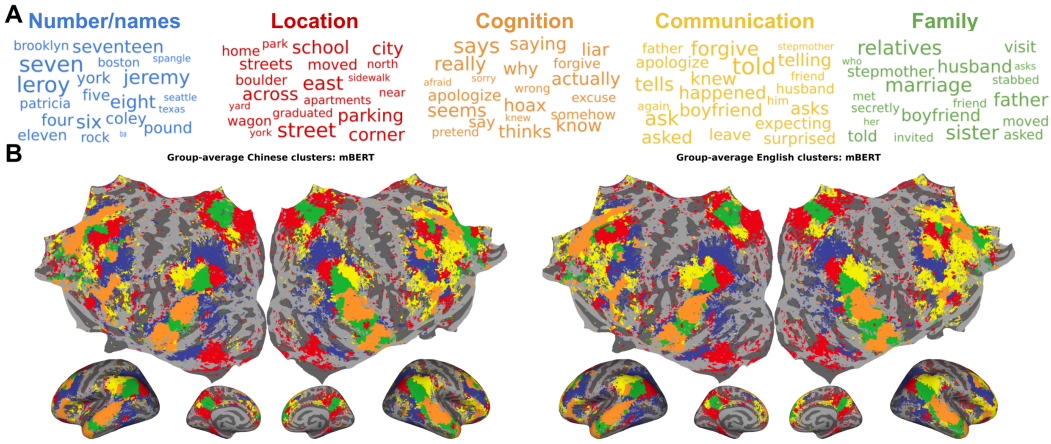

Figure 2: **Semantic clusters and their cortical location. A)** Semantic clusters were derived from the group-average encoding model weights. These clusters were interpreted by selecting the words closest to each cluster's centroid. **B)** Group-averaged semantic clusters are shown in each language on the common cortical surface. Vertices are colored according to the assigned cluster: number/names (blue), location (red), cognition (orange), communication (yellow), family (green). Vertices that are not well-predicted ($\sqrt{R^2} < 0.1$) are shown in gray. Semantic clusters span the cortex across the prefrontal, temporal and parietal cortices.

In order to interpret the semantic category associated with each cluster, we selected the closest English stimulus words to each cluster. We measured the distance between a word and a cluster as the Pearson correlation between the word's embedding and the cluster's centroid. Figure 2-A shows the 20 closest words to each cluster. The clusters were interpreted as number/names, location, cognition, communication and family.

Figure 2-B shows the cortical distribution of the group-averaged semantic clusters on the common brain space. Vertices are colored according to their assigned cluster: number/names in blue, location in red, cognition in orange, communication in yellow, family in green. Vertices that are not well-predicted at the group-average ($\sqrt{R^2_{zh}} < 0.1$ or $\sqrt{R^2_{en}} < 0.1$) are shown in gray. The clusters are distributed across the prefrontal, temporal and parietal cortices, with individual clusters spanning multiple spatially separated regions. Their distribution broadly reflects known functional specializations. For example, the location cluster (red) is located mainly in posterior areas in the temporal and parietal cortices, which is consistent with the role of these regions in representing places and scenes (Epstein & Kanwisher, 1998).

Voxels in each participant and language were then assigned to the cluster whose centroid is closest in Euclidean distance to the voxel's semantic model weights. To be able to compare cortical organization across participants, we projected the voxel weights and cluster assignments onto the common brain space. Individual participant maps in native and common brain space can be seen in Figure 6.

4.2 CORTICAL ORGANIZATION OF SEMANTIC CLUSTERS IS PARTICIPANT-SPECIFIC AND CONSISTENT ACROSS LANGUAGES

In order to compare the cortical organization of semantic representations across bilingual individuals, we measured the overlap of corresponding semantic clusters across participants and languages.

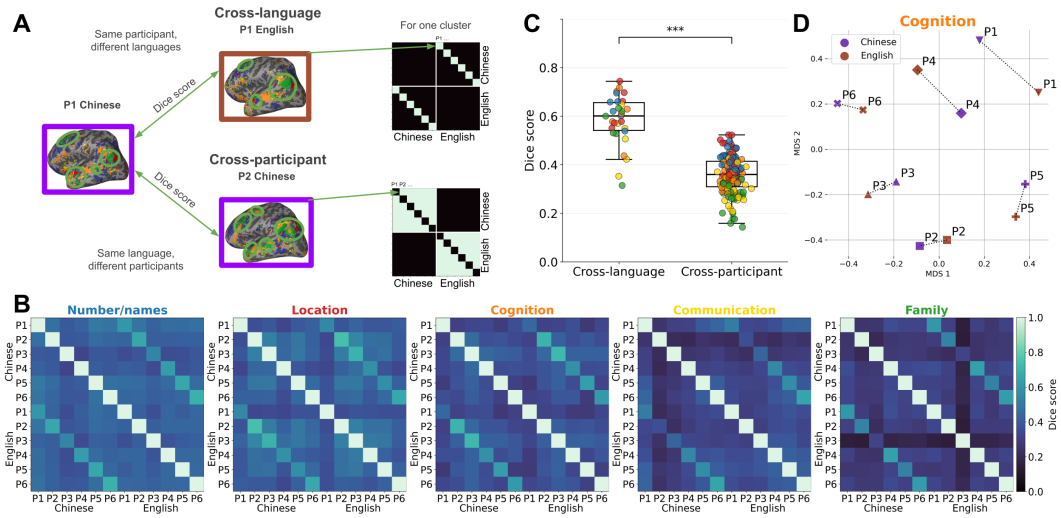

Figure 3: **Cortical overlap within each cluster. A)** For each cluster, cortical overlap was measured as the Dice score of vertices belonging to the given cluster across all participant-language pairs. **B)** Cortical overlap matrices for each of the five clusters. Rows and columns represent individual participant-language combinations, and cell color indicates the Dice score ranging from 0 (dark) to 1 (bright). **C)** Comparison of cross-language and cross-participant cortical overlap values. Cortical overlap is significantly higher across languages within one participant than across participants within one language ($p < 0.001$, one-sided paired t-test). **D)** 2D visualization of cortical overlap in MDS space for the cognition cluster (other clusters are shown in Figure 12).

Greater overlap indicates that vertices assigned to the same semantic cluster occupy more similar cortical locations. For each cluster, we computed cortical overlap as the Dice score of vertices assigned to that cluster across all participant-language pairs (Figure 3-A).

Figure 3-B shows the cortical overlap matrices for each cluster derived from mBERT weights (matrices for clusters derived from fastText weights are shown in Figure 8). Rows and columns are ordered by language, with the six participants shown first for Chinese and then for English. Each cell is colored according to the Dice score ranging from 0 in dark (no overlap at all) to 1 in bright (complete overlap). For all five semantic clusters, cortical overlap is high in the off-diagonals corresponding to cross-language comparisons within one participant. Cortical overlap is relatively low in the rest of the matrices. The greater within-participant overlap suggests that semantic clusters are organized in a participant-specific way that is shared across languages.

To quantify this effect, we compared cortical overlap across languages within one participant (*cross-language*) with cortical overlap across participants within one language (*cross-participant*). Figure 3-C shows the distribution of cortical overlap values for all pairs and clusters in each comparison. Cortical overlap is significantly higher for cross-language than cross-participant comparisons ($p < 0.001$, one-sided paired t-test). Figure 3-D visualizes this pattern for the cognition cluster (other clusters are shown in Figure 12-A). All participant-language combinations are represented in a two-dimensional multidimensional scaling (MDS) space, where proximity reflects greater cortical overlap. Each point represents one participant in one language (Chinese in purple and English in brown). Points corresponding to a given participant have the same shape and are connected by a dotted line. Their proximity illustrates the greater cortical overlap across languages within one participant than within language across participants.

4.3 Semantic tuning is largely shared across participants within one language

In order to compare semantic tuning across bilingual individuals, we measured the tuning similarity of corresponding semantic clusters across participants and languages. Greater tuning similarity

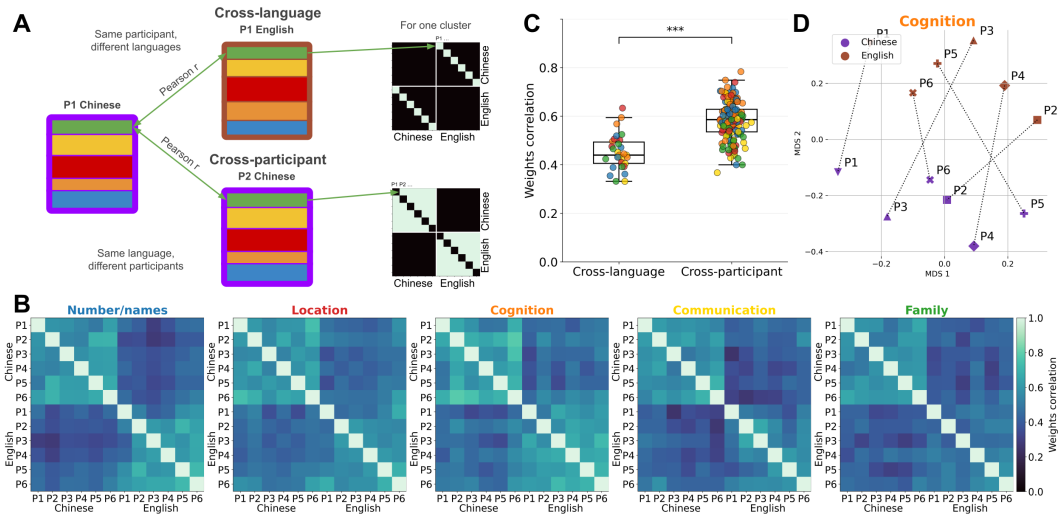

Figure 4: **Semantic tuning similarity within each cluster. A)** For each participant and language, model weights were averaged across vertices within each semantic cluster. Semantic tuning similarity was measured separately for each cluster as the Pearson correlation between these mean weight vectors across all participant-language pairs. **B)** Semantic tuning comparison matrices for all clusters. Rows and columns represent individual participant-language combinations, and cell color indicates weight correlations ranging from 0 (dark) to 1 (bright). **C)** Comparison of cross-language and cross-participant semantic tuning similarity values. Semantic tuning similarity is significantly higher across participants within one language than across languages within one participant ($p < 0.001$, one-sided paired t-test). **D)** 2D visualization of semantic tuning similarity in MDS space for the cognition cluster (other clusters are shown in Figure 12).

indicates that clusters are selective to more similar semantic features. For each participant and language, we averaged the estimated semantic model weights across vertices within each of the semantic clusters in order to obtain one mean semantic weight vector per cluster. For each cluster, we computed semantic tuning similarity as the Pearson correlation between the mean semantic weight vectors across all participant-language pairs (Figure 4-A). Figure 4-B shows the semantic tuning comparison matrices for each cluster derived from mBERT weights (matrices for clusters derived from fastText weights are shown in Figure 9). Rows and columns are ordered by language, with the six participants shown first for Chinese and then for English. Each cell is colored according to the Pearson correlation of the mean semantic weight vectors ranging from 0 (dark) to 1 (bright).

Qualitatively, for all five semantic clusters, semantic tuning similarity is high in the top-left block corresponding to cross-participant comparisons in Chinese and relatively high in the bottom-right block corresponding to cross-participant comparisons in English. Semantic tuning similarity is relatively low in the rest of the matrices. Clusters derived from fastText weights show consistent patterns, although cross-participant tuning similarity appears to be high in both Chinese and English (Figure 9). The greater cross-participant tuning similarity in each language suggests that tuning of semantic representations is largely shared across participants within one language.

To quantify this effect, we compared semantic tuning similarity across languages within one participant (*cross-language*) with semantic tuning similarity across participants within one language (*cross-participant*). Figure 4-C shows the distribution of semantic tuning similarity values for all pairs and clusters in each comparison. Semantic tuning similarity is significantly higher for cross-participant than cross-language comparisons ($p < 0.001$, one-sided paired t-test). Figure 4-D visualizes this pattern for the cognition cluster using multidimensional scaling (MDS; other clusters are shown in Figure 12-C). All participant-language combinations are represented in the two-dimensional MDS space, where proximity reflects greater semantic tuning similarity. Each point represents one participant in one language (Chinese in purple and English in brown). Points corresponding to a given participant have the same shape and are connected by a dotted line. The

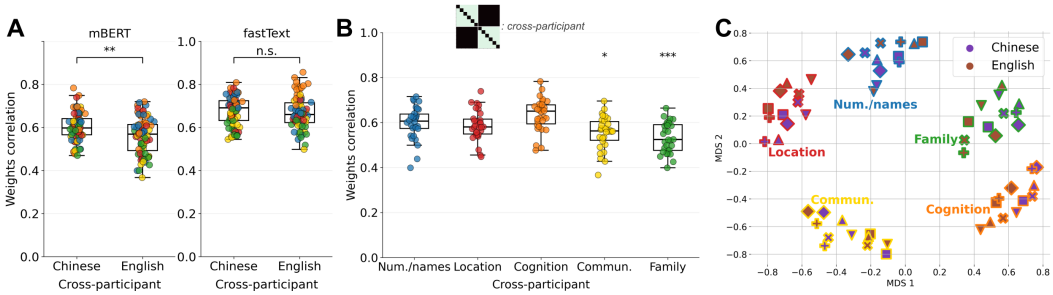

Figure 5: **Cross-participant semantic tuning similarity across languages and semantic clusters.** In A) and B), each point represents the weight correlation across two participants within one language and cluster. **A)** Cross-participant semantic tuning similarity in English and in Chinese, shown separately for mBERT and fastText weights. Similarity is significantly higher in Chinese than in English for mBERT weights ($p = 0.002$, one-sided paired t-test) but does not differ significantly between languages for fastText weights ($p = 0.203$, one-sided paired t-test). **B)** Cross-participant semantic tuning similarity for each semantic cluster. Similarity is significantly lower for the communication and family clusters than the mean of the others ($p = 0.027$, $p < 0.001$, Bonferroni-corrected one-sided paired t-tests). This suggests that semantic representations of these categories are more participant-specific. **C)** Two-dimensional MDS visualization of semantic tuning similarity across participants, languages and clusters. Points are grouped primarily by semantic cluster and, within each cluster, are approximately separated by language.

separation of points by language illustrates the greater similarity of semantic tuning across participants within one language than within participant across languages.

4.4 Semantic tuning differences between native and non-native languages

Visual inspection of semantic tuning comparison matrices suggests that cross-participant similarity is higher in Chinese than in English for clusters derived from mBERT weights (Figure 4-B), but is similarly high in the two languages for clusters derived from fastText weights (Figure 9). To verify this observation, we compared cross-participant semantic tuning similarity between Chinese and English separately for clusters derived from mBERT and fastText weights. Figure 5-A shows the distributions of these values for each language and embedding space.

For clusters derived from mBERT weights, cross-participant semantic tuning similarity is significantly higher in Chinese than in English ($p = 0.002$, one-sided paired t-test). There is no significant difference between the two languages for clusters derived from fastText weights ($p = 0.203$, one-sided paired t-test). All participants were native Chinese speakers who acquired English as a second language. This may suggest that differences in cross-participant semantic tuning similarity between native and non-native languages could depend on the aspects of semantic tuning captured by each model. Greater similarity in the native language may be more apparent for aspects captured by contextual representations than for those captured by lexical representations.

4.5 Semantic representations of social concepts are more participant-specific

In order to assess whether semantic representations vary more across participants for some semantic categories than others, we compared cross-participant semantic tuning similarity for each cluster with the average across the remaining clusters. Figure 5-B shows the distribution of these values for each cluster derived from mBERT weights. We compared each cluster's similarity values with the corresponding mean across the other four clusters, pairing values by participant pair and language. The communication and family clusters were the only clusters that showed significantly lower cross-participant semantic tuning similarity than the average of the other clusters ($p = 0.027$, $p < 0.001$, respectively, one-sided paired t-tests, Bonferroni-corrected across five comparisons). This pattern is also observed for cortical overlap and for clusters derived from fastText weights (Figure 11). This suggests that the semantic representations of communication and family concepts, both involving social content, are more participant-specific.

To examine how these differences relate to the overall pattern of semantic tuning similarity, we extended the similarity analysis to include comparisons across clusters. Specifically, we computed semantic tuning similarity as the Pearson correlation of the mean semantic weight vectors across participants, languages and clusters. The resulting 60×60 comparison matrix is shown in Figure 13. Figure 5-C visualizes semantic tuning similarity across all participant-language-cluster combinations in a two-dimensional MDS space. Each point represents one participant in one language and one cluster. The points' shape indicates the participants, the fill color indicates the language (Chinese in purple and English in brown) and the outline color indicates the cluster. Points are grouped primarily by semantic cluster and, within each cluster, they are approximately separated by language. This organization indicates that the clusters have distinct semantic tuning profiles at a coarse level, while the fine-grained similarity across participants within the same language is consistent across clusters.

5 Conclusion and discussion

In this study, we examined how semantic representations vary across participants and languages in bilingual brains. We compared semantic clusters derived from six Chinese-English bilingual participants in terms of spatial cortical organization and semantic tuning.

First, we show that for all semantic clusters, cortical overlap is higher across languages within one participant than across participants (Figure 3). This suggests that cortical organization of semantic representations is participant-specific and consistent across languages within an individual. Second, we find the opposite pattern for semantic tuning. For all semantic clusters as well, semantic tuning similarity is higher across participants within one language than across languages within one participant (Figure 4). This suggests that semantic tuning is largely shared across participants, but differs between languages even within an individual, consistent with the fine-grained language-specific differences reported by Chen et al. (2026).

The comparison between clusters derived from mBERT and fastText raises the possibility that differences between native and non-native language depend on the aspects of semantic tuning captured by each model. With mBERT, semantic tuning appears more similar across participants in their native than in the non-native languages, whereas there is no significant difference with fastText (Figure 5-A). One possible interpretation is that contextual representations may capture deeper aspects of semantic representations which are more sensitive to language nativeness (Li et al., 2023). However, this interpretation remains tentative and comparisons with additional contextual and lexical models are needed to assess its generality. Cross-participant similarity also differed across semantic categories. Clusters associated with social concepts, such as family and communication, showed lower cross-participant semantic tuning similarity (Figure 5-B) and cortical overlap (Figure 11). This pattern is consistent with previous findings of greater individual variability for social than for physical concepts (Visconti di Oleggio Castello et al., 2025), and suggests that such variability concerns both semantic tuning and cortical organization.

Overall, our findings highlight the importance of distinguishing where semantic representations are located from how they are tuned. Across bilingual individuals, semantic tuning is largely shared with fine-grained language-specific differences, while spatial cortical organization is participant-specific and preserved across languages.

6 Limitations and future work

Comparisons of cortical organization across participants rely on mapping individual cortical surfaces to the common fsaverage space. Imperfect alignment may reduce the measured cortical overlap across participants. Future work could examine how robust this pattern is to different alignment methods.

Our sample includes only six Chinese-English bilingual participants. Although extensive data were collected from each participant in each language, the small sample limits the generalizability of our findings. Future studies with larger and more diverse samples could examine whether these patterns extend to other language pairs and how they vary with language proficiency and age of acquisition.

AI USE STATEMENT

In this work, we have not used generative AI tools for helping develop theoretical models or conceptual frameworks, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, implementing methods, assisting with translation, cleaning and reformating dataset, supporting qualitative and thematic data analysis, interpreting results. Generating of synthetic data sets, formulating mathematical claims, providing critical ingredients for proving mathematical claims and assisting in the writing of proofs are not applicable to this work.

Additionally, we used generative AI tools for editing the research paper to improve readability and editing software code to modify figures. We have reviewed all AI-assisted work. We manually checked LLM-generated suggestions to the writing of paragraphs and all the edits to the software code used to modify figures.

We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This study uses previously collected human neuroimaging data (Chen et al., 2026). All procedures in the original study were approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley, and all participants gave informed consent.

REPRODUCIBILITY STATEMENT

The data used in this study are publicly available (Chen et al., 2026). Details of the preprocessing, model fitting and analyses are provided in the Methodology Section. The code used to reproduce the analyses and figures will be made publicly available upon acceptance.

REFERENCES

Jubin Abutalebi and David Green. Bilingual language production: The neurocognition of language representation and control. *Journal of neurolinguistics*, 20(3):242–275, 2007.

Jubin Abutalebi, Stefano F Cappa, and Daniela Perani. The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and cognition*, 4(2):179–190, 2001.

Edward H Adelson and James R Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the optical society of america A*, 2(2):284–299, 1985.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146, 2017. ISSN 2307-387X.

Thomas L Botch and Emily S Finn. Neural representations of concreteness and concrete concepts are specific to the individual. *The Journal of Neuroscience*, 44(45):e0288242024, 2024.

Augusto Buchweitz, Svetlana V Shinkareva, Robert A Mason, Tom M Mitchell, and Marcel Adam Just. Identifying bilingual semantic neural representations across languages. *Brain and language*, 120(3):282–289, 2012.

Charlotte Caucheteux and Jean-Rémi King. Brains and algorithms partially converge in natural language processing. *Communications biology*, 5(1):134, 2022.

Catherine Chen, Tom Dupré la Tour, Jack L Gallant, Daniel Klein, and Fatma Deniz. The cortical representation of language timescales is shared between reading and listening. *Communications Biology*, 7(1):284, 2024.

Catherine Chen, Xue L Gong, Christine Tseng, Daniel L Klein, Jack L Gallant, and Fatma Deniz. Bilingual language processing relies on shared semantic representations that are modulated by each language. *Proceedings of the National Academy of Sciences*, 123(9):e2503721123, 2026.

Stanislas Dehaene, Emmanuel Dupoux, Jacques Mehler, Laurent Cohen, Eraldo Paulesu, Daniela Perani, Pierre-Francois Van de Moortele, Stéphane Lehéricy, and Denis Le Bihan. Anatomical variability in the cortical representation of first and second language. *Neuroreport*, 8(17):3809–3815, 1997.

Morteza Dehghani, Reihane Boghrati, Kingson Man, Joe Hoover, Sarah I Gimbel, Ashish Vaswani, Jason D Zevin, Mary Helen Immordino-Yang, Andrew S Gordon, Antonio Damasio, et al. Decoding the neural representation of story meanings across languages. *Human brain mapping*, 38 (12):6096–6106, 2017.

Fatma Deniz, Anwar O Nunez-Elizalde, Alexander G Huth, and Jack L Gallant. The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *The journal of neuroscience*, 39(39):7722–7736, 2019.

Fatma Deniz, Christine Tseng, Leila Wehbe, Tom Dupré la Tour, and Jack L Gallant. Semantic representations during language comprehension are affected by context. *The Journal of Neuroscience*, 43(17):3144–3158, 2023.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 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, volume 1 (long and short papers)*, pp. 4171–4186, 2019.

Russell Epstein and Nancy Kanwisher. A cortical representation of the local visual environment. *Nature*, 392(6676):598–601, 1998.

Evelina Fedorenko and Nancy Kanwisher. Neuroimaging of language: why hasn't a clearer picture emerged? *Language and linguistics compass*, 3(4):839–865, 2009.

Bruce Fischl, Martin I Sereno, Roger BH Tootell, and Anders M Dale. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human brain mapping*, 8(4):272–284, 1999.

Xue Gong. *Language representation in human cerebral cortex*. PhD thesis, UC Berkeley, 2024.

Xue L Gong, Alexander G Huth, Fatma Deniz, Keith Johnson, Jack L Gallant, and Frédéric E Theunissen. Phonemic segmentation of narrative speech in human cerebral cortex. *Nature communications*, 14(1):4309, 2023.

Christopher J Honey, Christopher R Thompson, Yulia Lerner, and Uri Hasson. Not lost in translation: neural responses shared across languages. *The Journal of Neuroscience*, 32(44):15277–15283, 2012.

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

Judy Illes, Wendy S Francis, John E Desmond, John DE Gabrieli, Gary H Glover, Russell Poldrack, Christine J Lee, and Anthony D Wagner. Convergent cortical representation of semantic processing in bilinguals. *Brain and language*, 70(3):347–363, 1999.

Armand Joulin, Piotr Bojanowski, Tomas Mikolov, Hervé Jégou, and Edouard Grave. Loss in translation: Learning bilingual word mapping with a retrieval criterion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018.

Meenakshi Khosla, Gia H Ngo, Keith Jamison, Amy Kuceyeski, and Mert R Sabuncu. Cortical response to naturalistic stimuli is largely predictable with deep neural networks. *Science Advances*, 7(22):eabe7547, 2021.

Karl HS Kim, Norman R Relkin, Kyoung-Min Lee, and Joy Hirsch. Distinct cortical areas associated with native and second languages. *Nature*, 388(6638):171–174, 1997.

Denise Klein, Brenda Milner, Robert J Zatorre, Ernst Meyer, and Alan C Evans. The neural substrates underlying word generation: a bilingual functional-imaging study. *Proceedings of the National Academy of Sciences*, 92(7):2899–2903, 1995.

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

Mathis Lamarre, Catherine Chen, and Fatma Deniz. Attention weights accurately predict language representations in the brain. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pp. 4513–4529, 2022.

Mathis Lamarre, Xue L Gong, Catherine Chen, and Fatma Deniz. Concreteness shapes semantic representations in bilingual brains. *bioRxiv*, pp. 2025–12, 2025.

Huiling Li, Ying Cao, Chuansheng Chen, Xiaoyu Liu, Shuo Zhang, and Leilei Mei. The depth of semantic processing modulates cross-language pattern similarity in chinese–english bilinguals. *Human brain mapping*, 44(5):2085–2098, 2023.

Kang-Kwong Luke, Ho-Ling Liu, Yo-Yo Wai, Yung-Liang Wan, and Li Hai Tan. Functional anatomy of syntactic and semantic processing in language comprehension. *Human brain mapping*, 16(3):133–145, 2002.

Saima Malik-Moraleda, Dima Ayyash, Jeanne Gallée, Josef Affourtit, Malte Hoffmann, Zachary Mineroff, Olessia Jouravlev, and Evelina Fedorenko. An investigation across 45 languages and 12 language families reveals a universal language network. *Nature neuroscience*, 25(8):1014–1019, 2022.

Emily X Meschke, Matteo Visconti di Oleggio Castello, Tom Dupré la Tour, and Jack L Gallant. Model connectivity: leveraging the power of encoding models to overcome the limitations of functional connectivity. *BioRxiv*, pp. 2023–07, 2023.

Tomoya Nakai, Hiroto Q Yamaguchi, and Shinji Nishimoto. Convergence of modality invariance and attention selectivity in the cortical semantic circuit. *Cerebral Cortex*, 31(10):4825–4839, 2021.

Thomas Naselaris, Kendrick N Kay, Shinji Nishimoto, and Jack L Gallant. Encoding and decoding in fmri. *Neuroimage*, 56(2):400–410, 2011.

Anuja Negi, Christine Tseng, Anwar O Nunez-Elizalde, Xue Lily Gong, and Fatma Deniz. Optimizing language model embeddings to voxel activity improves brain activity predictions. *bioRxiv*, pp. 2025–09, 2025.

Shinji Nishimoto and Jack L Gallant. A three-dimensional spatiotemporal receptive field model explains responses of area mt neurons to naturalistic movies. *Journal of Neuroscience*, 31(41): 14551–14564, 2011.

Anwar O Nunez-Elizalde, Alexander G Huth, and Jack L Gallant. Voxelwise encoding models with non-spherical multivariate normal priors. *Neuroimage*, 197:482–492, 2019.

AO Nunez-Elizalde, F Deniz, T Dupré la Tour, M Visconti di Oleggio Castello, and JL Gallant. Pymoten: scientific python package for computing motion energy features from video. *Zenodo doi*, 10, 2021.

Daniela Perani and Jubin Abutalebi. The neural basis of first and second language processing. *Current opinion in neurobiology*, 15(2):202–206, 2005.

Sara F Popham, Alexander G Huth, Natalia Y Bilenko, Fatma Deniz, James S Gao, Anwar O Nunez-Elizalde, and Jack L Gallant. Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nature neuroscience*, 24(11):1628–1636, 2021.

Shirley-Ann Rüschemeyer, Stefan Zysset, and Angela D Friederici. Native and non-native reading of sentences: An fmri experiment. *NeuroImage*, 31(1):354–365, 2006.

Melissa Thye, Paul Hoffman, and Daniel Mirman. The neural basis of naturalistic semantic and social cognition. *Scientific Reports*, 14(1):6796, 2024.

Mariya Toneva and Leila Wehbe. Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). *Advances in neural information processing systems*, 32, 2019.

Matteo Visconti di Oleggio Castello, Tom Dupré la Tour, and Jack L Gallant. Individual differences shape conceptual representation in the brain. *bioRxiv*, pp. 2025–08, 2025.

Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

Andrew B Watson and Albert J Ahumada Jr. Model of human visual-motion sensing. *Journal of the optical Society of America A*, 2(2):322–342, 1985.

Stephen M Wilson, Melodie Yen, and Dana K Eriksson. A n adaptive semantic matching paradigm for reliable and valid language mapping in individuals with aphasia. *Human brain mapping*, 39 (8):3285–3307, 2018.

Michael C-K Wu, Stephen V David, and Jack L Gallant. Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.*, 29(1):477–505, 2006.

Min Xu, Daniel Baldauf, Chun Qi Chang, Robert Desimone, and Li Hai Tan. Distinct distributed patterns of neural activity are associated with two languages in the bilingual brain. *Science advances*, 3(7):e1603309, 2017.

Ying Yang, Jing Wang, Cyntia Bailer, Vladimir Cherkassky, and Marcel Adam Just. Commonalities and differences in the neural representations of english, portuguese, and mandarin sentences: When knowledge of the brain-language mappings for two languages is better than one. *Brain and language*, 175:77–85, 2017a.

Ying Yang, Jing Wang, Cyntia Bailer, Vladimir Cherkassky, and Marcel Adam Just. Commonality of neural representations of sentences across languages: Predicting brain activation during portuguese sentence comprehension using an english-based model of brain function. *Neuroimage*, 146:658–666, 2017b.

Satoru Yokoyama, Hideyuki Okamoto, Tadao Miyamoto, Kei Yoshimoto, Jungho Kim, Kazuki Iwata, Hyeonjeong Jeong, Shinya Uchida, Naho Ikuta, Yuko Sassa, et al. Cortical activation in the processing of passive sentences in l1 and l2: An fmri study. *Neuroimage*, 30(2):570–579, 2006.

A Appendix

A.1 Cortical maps of all participants

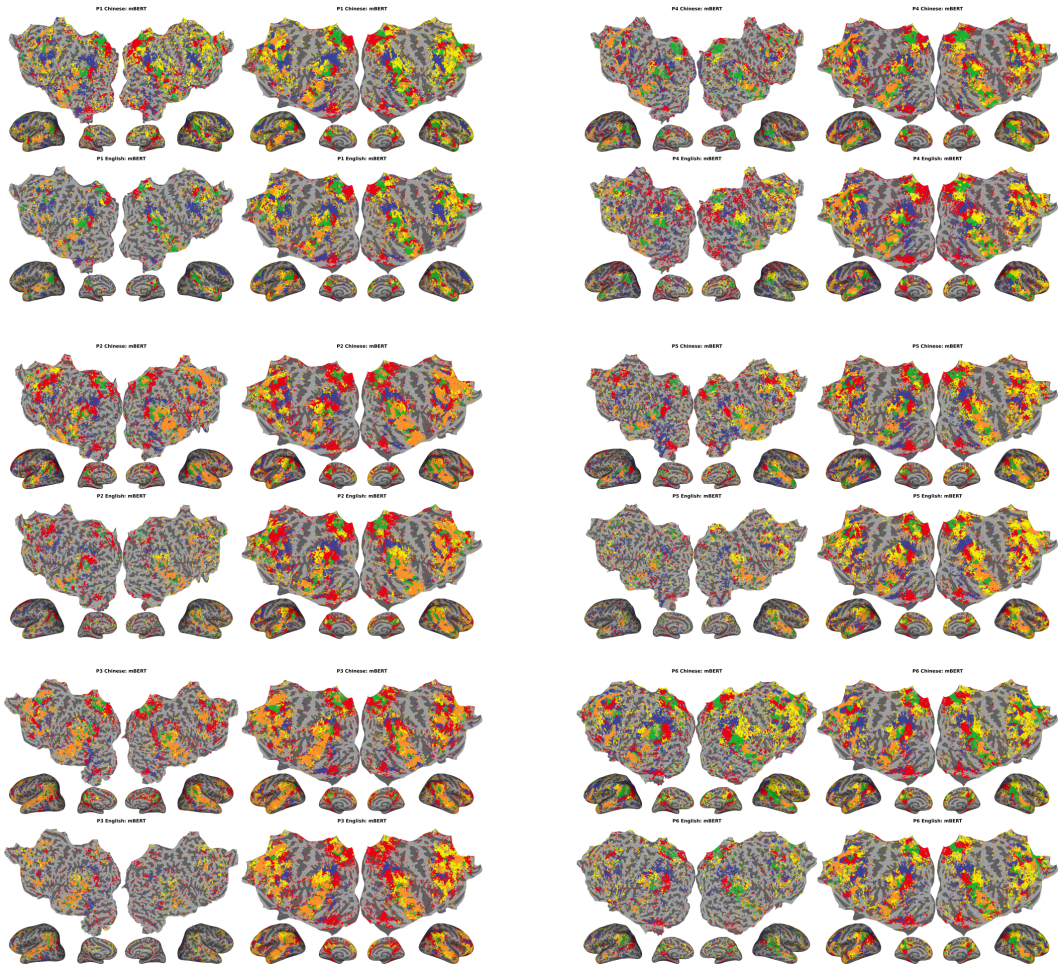

Figure 6: **Semantic clusters derived from mBERT weights on the cortical surface of all participants.** Semantic clusters are shown for both languages for each participant on their native (left) and common (right) cortical surface. Vertices are colored according to the assigned cluster: number/names (blue), location (red), cognition (orange), communication (yellow), family (green). Vertices that are not well-predicted ($\sqrt{R^2} < 0.1$) are shown in gray.

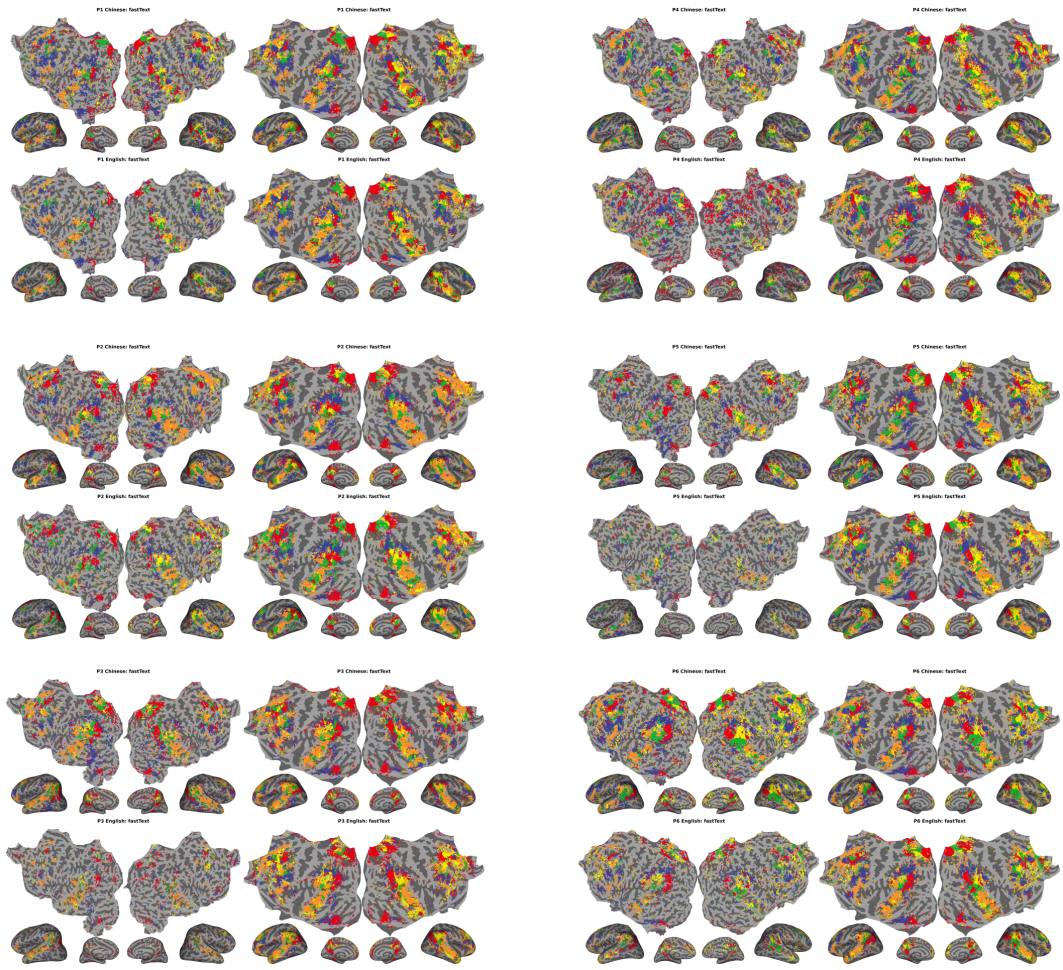

Figure 7: **Semantic clusters derived from fastText weights on the cortical surface of all participants.** Semantic clusters are shown for both languages for each participant on their native (left) and common (right) cortical surface. Vertices are colored according to the assigned cluster: family (green), communication (yellow), cognition (orange), location (red), number/names (blue). Vertices that are not well-predicted ($\sqrt{R^2} < 0.1$) are shown in gray.

A.2 Other cortical organization and semantic tuning comparisons

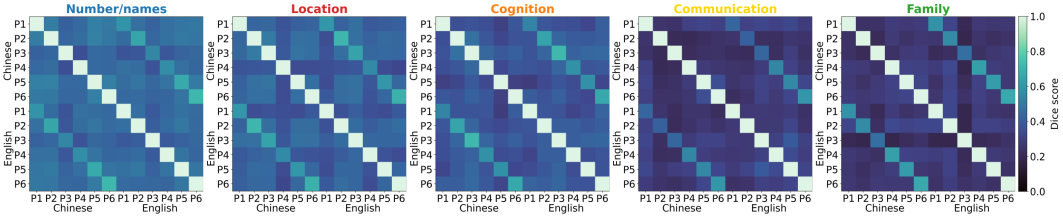

Figure 8: **Cortical overlap for each of the five clusters derived from fastText weights**. Rows and columns represent individual participant-language combinations, and cell color indicates the Dice score ranging from 0 (dark) to 1 (bright). In all clusters, overlap is higher in the off-diagonals than in the rest of the matrix. This suggests that cortical organization is participant-specific and largely shared across languages.

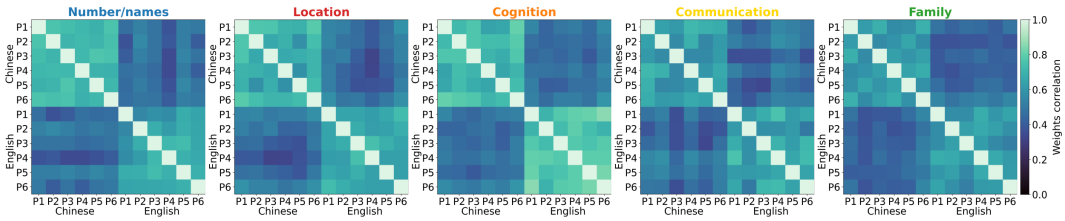

Figure 9: **Semantic tuning similarity for each of the five clusters derived from fastText weights**. Rows and columns represent individual participant-language combinations, and cell color indicates the weights correlation ranging from 0 (dark) to 1 (bright). In all clusters, tuning similarity is higher in the top-left and bottom-right blocks than in the rest of the matrices. This suggests that semantic tuning is largely shared across participants within one language.

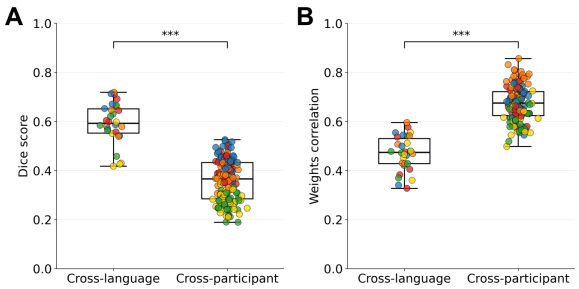

Figure 10: **Cross-language and cross-participant cortical overlap and semantic tuning similarity values with fastText weights.** Each cross-language point represents a comparison across languages within one participant. Each cross-participant point represents a comparison across two participants within one language. **A)** Cortical overlap is significantly higher across languages than across participants ($p < 0.001$, one-sided paired t-test). **B)** Semantic tuning similarity is significantly higher across participants than across languages ($p < 0.001$, one-sided paired t-test).

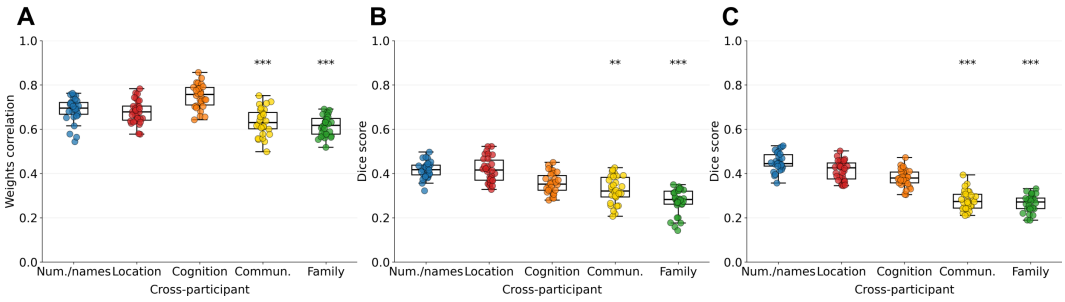

Figure 11: **Cluster-wise cross-participant semantic tuning similarity and cortical overlap with mBERT and fastText weights.** Each point represents a comparison across two participants within one language, and values are grouped by cluster. We compared the values in each cluster to the average of the other four clusters (one-sided paired t-tests Bonferroni-corrected for multiple comparisons). Values are significantly lower only for the communication and family clusters. **A)** Semantic tuning similarity with clusters derived from fastText weights ($p < 0.001$, $p < 0.001$, respectively for communication and family). **B)** Cortical overlap with clusters derived from mBERT weights ($p = 0.004$, $p < 0.001$, respectively for communication and family). **C)** Cortical overlap with clusters derived from fastText weights ($p < 0.001$, $p < 0.001$, respectively for communication and family). This suggests that representations of communication and family are more participant-specific.

A.3 MDS VISUALIZATIONS FOR ALL CLUSTERS

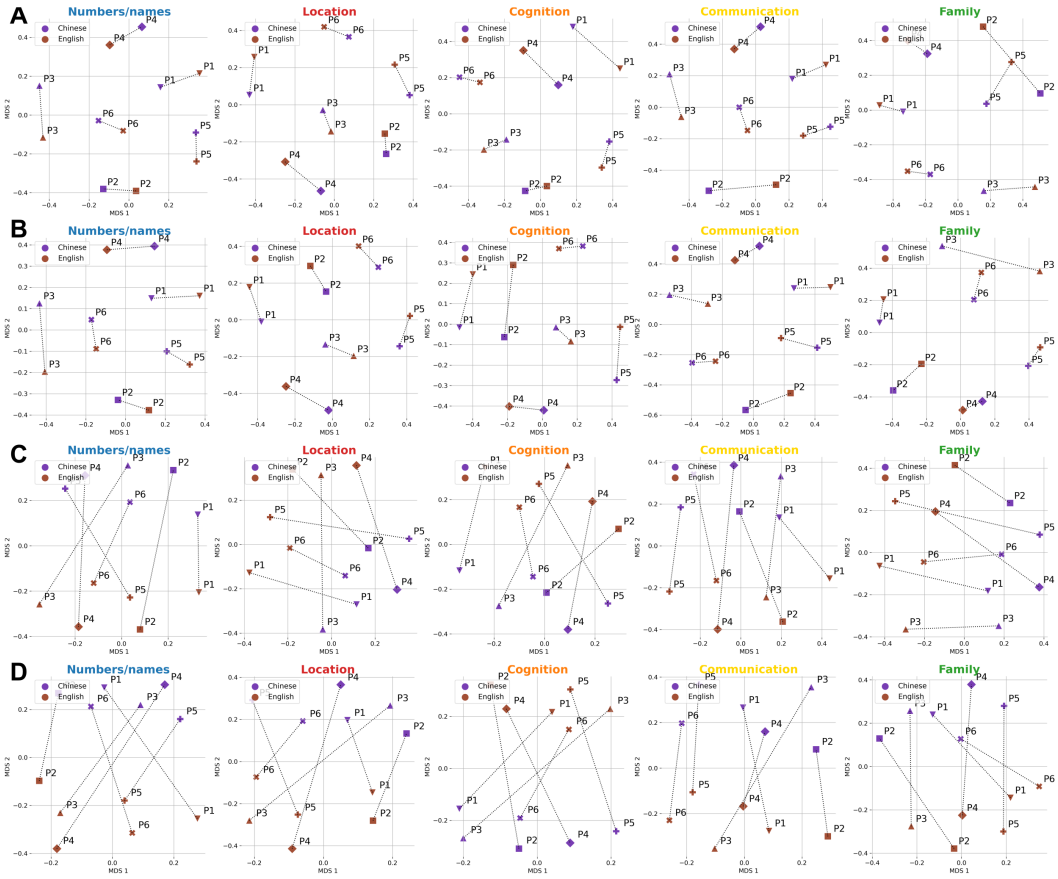

Figure 12: **Two-dimensional MDS space visualizations of cortical overlap and semantic tuning similarity for all clusters.** Each point represents one participant in one language, with Chinese shown in purple and English in brown. Points corresponding to the same participant are connected by a dotted line. Cortical overlap for clusters derived from mBERT weights **(A)** and for clusters derived from fastText weights **(B)**. The proximity of points corresponding to the same participant illustrates the greater cortical overlap across languages within one participant than across participants. Semantic tuning similarity for clusters derived from mBERT weights **(C)** and for clusters derived from fastText weights **(D)**. The separation of points by language illustrates the greater similarity of semantic tuning across participants within one language than across languages.

A.4 CROSS-CLUSTER SEMANTIC TUNING COMPARISONS

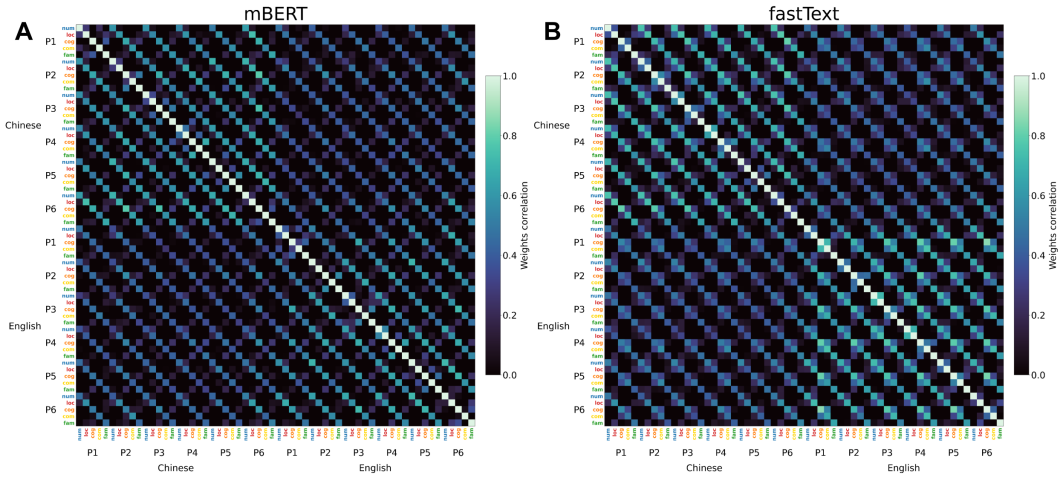

Figure 13: **Cross-cluster semantic tuning similarity.** Semantic tuning similarity across participants, languages and clusters shown for clusters derived from mBERT weights (**A**) and for clusters derived from fastText weights (**B**).

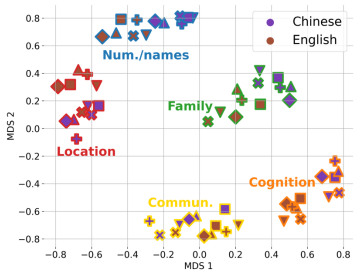

Figure 14: **Two-dimensional MDS visualization of cross-cluster semantic tuning similarity with clusters derived from fastText weights.** Each point represents one participant in one language and one cluster, with Chinese shown in purple and English in brown. Points are grouped primarily by semantic cluster and, within each cluster, are approximately separated by language. This suggests that the clusters have distinct semantic tuning profiles, while the pattern of greater similarity across participants within the same language is broadly consistent across clusters.