Language-Neutral or Language-Specific? Disentangling Multilingual LLM Subspaces with Naturalistic fMRI

Ali Uyar
Independent Researcher

Abstract

Large language model features increasingly explain naturalistic fMRI responses, but most model–brain alignment studies remain effectively monolingual. Multilingual encoders, by contrast, contain both cross-lingually shared structure and language-specific residual structure, which makes them useful probes for a more mechanistic question: during comprehension, is the brain better predicted by what several languages agree on or by the information that remains specific to the heard language? We tested this question in the *Le Petit Prince* multilingual naturalistic fMRI corpus by aligning English, French, and Chinese sentence spans, extracting sentence representations from XLM-R-base and the encoder of NLLB-200-distilled-600M, decomposing each target-language representation into a leave-target-out shared component and an orthogonalized target-language-specific residual, and evaluating those components with cross-validated ROI encoding models. In the semantic ROI family, the shared component outperformed the specific residual in all six model-by-language confirmatory tests, and every H1 effect survived Holm correction. The stronger prediction that the semantic-family shared advantage would exceed the auditory-family shared advantage remained mixed, because auditory and superior temporal ROIs also showed robust `SHARED`–`SPECIFIC` effects. Correct-content controls remained directionally healthy, including broad `SHARED` > `MISMATCHED_SHARED` behavior, whereas text-geometry-to-brain coupling was weak and non-significant in the present checkpoint. Robustness analyses supported the main semantic-family effect under several perturbations, but a voxelwise-within-ROI target produced negative French rows in both model families. The most defensible conclusion is therefore narrower than a clean semantic-versus-auditory dissociation: multilingual LLMs can serve as factorized probes of naturalistic language processing, but the brain-side distribution of shared and specific sensitivity is regionally heterogeneous.

Keywords: multilingual language models; naturalistic fMRI; language neuroscience; encoding models; cross-lingual representations

1 Introduction

Encoding models built from contemporary language representations have become a productive way to ask which aspects of linguistic computation covary with naturalistic brain responses (Toneva and Wehbe, 2019; Schrimpf et al., 2021; Caucheteux and King, 2022). In practice, however, most of this literature is still organized one language at a time. That design choice obscures a question that multilingual models make explicit: when the same content is expressed in several languages, does

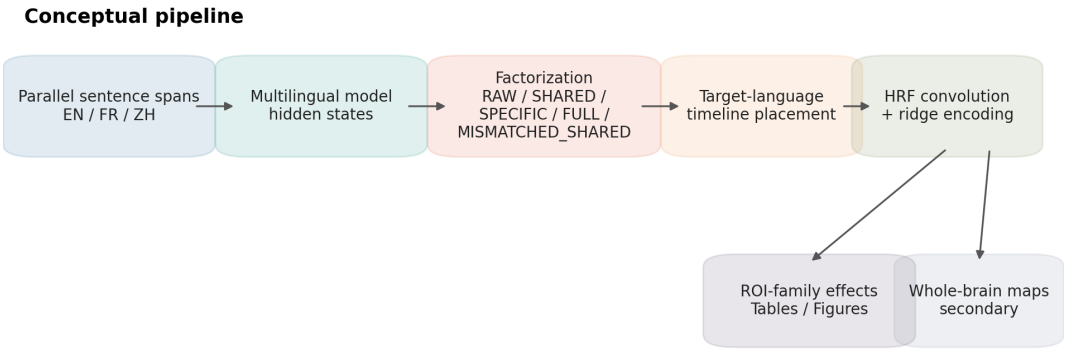

Figure 1: Conceptual overview of the multilingual encoding pipeline. Parallel sentence spans in English, French, and Chinese are encoded by multilingual language models, decomposed into a leave-target-out `SHARED` component and an orthogonalized target-language-specific `SPECIFIC` residual, placed on the target listener's own timeline, convolved with a canonical hemodynamic response function, and evaluated with cross-validated ROI encoding models. The critical operational point is that `SHARED` is built from the other languages but aligned to the target subject's own sentence timing.

the explanatory signal for brain responses come primarily from a cross-lingually shared component, or from the residual structure that remains specific to the heard language?

That question matters because multilingual encoders are neither fully language-invariant nor merely monolingual models with a larger vocabulary. Their sentence spaces contain both convergence across languages and residual language-sensitive structure. If these components can be separated directly, then multilingual models become more than feature extractors: they become factorized probes for testing what kind of information naturalistic language cortex tracks.

Naturalistic multilingual fMRI provides a particularly strong setting for this test. The same narrative can be presented in different languages while preserving broad content structure, and sentence-level timing makes it possible to compare content-matched but language-different representations against target-language BOLD responses on the listener's own time axis. The *Le Petit Prince* multilingual naturalistic fMRI corpus (LPPC) is especially well suited to this problem because it combines a shared story stimulus, public derivatives, and open annotations across English, French, and Chinese (Li et al., 2022). What LPPC does not provide by itself is a mechanistic decomposition that distinguishes a leave-target-out shared signal from a target-language-specific residual and tests both against the same neural targets.

We therefore ask a focused question: in multilingual naturalistic comprehension, does a leave-target-out shared representation predict language-network activity better than an orthogonalized target-language-specific residual? Our contribution is fourfold. First, we build a reproducible multilingual sentence-span and triplet-alignment pipeline with canonical run remapping and explicit tri-lingual quality control. Second, we decompose multilingual sentence embeddings into a shared component and an orthogonalized specific residual. Third, we evaluate those components in ROI-first naturalistic encoding models with required controls, including `RAW`, `FULL`, `MISMATCHED_SHARED`, and an acoustic-only baseline. Fourth, we compare text-side multilingual geometry to brain-side shared advantage while keeping confirmatory and descriptive analyses sharply separated.

2 Related Work

Model–brain alignment studies have repeatedly shown that distributed language representations explain variance in naturalistic cortical responses, especially in higher-level temporal and frontal regions (Toneva and Wehbe, 2019; Schrimpf et al., 2021; Caucheteux and King, 2022). For the present problem, the main limitation is not whether language models carry explanatory signal, but that most analyses remain monolingual and therefore cannot ask whether the predictive component is genuinely cross-lingually shared.

Multilingual representation learning provides the right computational contrast. XLM-R was designed as a multilingual masked-language encoder supporting large-scale cross-lingual transfer, while NLLB was trained to preserve rich multilingual structure for machine translation across hundreds of languages (Conneau et al., 2020; NLLB Team et al., 2024). These models encourage convergence across languages without forcing complete invariance, which makes them plausible candidates for a shared-versus-specific factorization.

On the neuroscience side, LPPC offers a rare multilingual naturalistic fMRI resource built around the same narrative in English, French, and Chinese (Li et al., 2022). More broadly, naturalistic neuroimaging datasets have shown the value of story-based paradigms for evaluating computational models of language comprehension (Nastase et al., 2021). LPPC therefore enables a more mechanistic multilingual question than unmatched sentence tasks or separate stimuli could support.

Methodologically, our work sits within standard naturalistic encoding practice: nuisance regression fit on training data only, dimensionality reduction fit on training data only, regularized linear prediction, and held-out evaluation in Fisher-z space. The novelty is not the regression family itself, but the explicit factorization of multilingual sentence representations and the way that factorization is tested against brain activity (Abraham et al., 2014).

3 Materials and Methods

3.1 Dataset and analysis sample

The core dataset was LPPC (OpenNeuro `ds003643`) (Li et al., 2022). Derivative discovery identified 112 participants in total (49 English, 28 French, and 35 Chinese) and 1008 run-level derivative rows in the canonical run manifest. The frozen ROI analysis bundle used for the present manuscript contained 94 participants with complete ROI targets: 31 English, 28 French, and 35 Chinese listeners, each with all nine canonical runs. Mean story duration was 93.87 minutes in English, 97.90 minutes in French, and 99.23 minutes in Chinese, with TR = 2.0 s (Table 1).

We used the public preprocessed derivatives as distributed, rather than reprocessing from raw MRI. These derivatives are expressed in LPPC's `MNIColin27` space, which creates a template mismatch relative to standard `MNI152`-distributed atlas products. We therefore treat anatomical precision conservatively and frame the paper as ROI-first from the outset. Because the three language cohorts were collected at different sites and scanners, we do not interpret cross-language magnitude differences as scanner-free quantities.

Table 1: LPPC analysis sample used in the core manuscript. All cohorts include nine canonical runs and public story annotations.

Language	Subjects	Runs	Duration (min)	TR (s)
EN	31	9	93.87	2.0
FR	28	9	97.90	2.0
ZH	35	9	99.23	2.0

The broader derivative inventory contains 112 participants and 1008 discovered run rows; the frozen ROI bundle used here contains the 94 participants with complete ROI targets.

3.2 Canonical run mapping, sentence spans, and multilingual triplets

LPPC derivative filenames can use nonconsecutive raw run labels rather than the story's canonical run indices. We therefore discovered all run files per subject, sorted them by original run label, and remapped them to canonical runs 1–9 in a saved manifest rather than assuming that the filename run number already matched the story section.

Sentence spans were the primary unit of analysis. We built them from the public LPPC annotations by matching sentence structure to language-specific word-timing files. Each span inherited its onset from the first matched word onset and its offset from the final matched word offset. When alignment required a two-sentence merge, we concatenated the texts with a single space and inherited the onset and offset from the outer sentence boundaries; larger merges were not permitted in the core pipeline.

Multilingual triplets were aligned within matched story sections and canonical runs, using English as the implementation pivot. Pairwise English–French and English–Chinese alignments were obtained with a monotonic dynamic-programming procedure combining LaBSE-based semantic similarity, a character-ratio penalty, punctuation mismatch penalties, and a small merge cost. Conflicting pairwise windows were then repaired locally with a tri-lingual reconciliation pass rather than allowing one pairwise path to dominate. The final alignment table contained 1323 triplets, with 284 rows flagged for review, 206 conflict windows repaired, and one logged unresolved outlier in section 9. Most rows were strict 1–1–1 alignments, but the final table also retained admissible two-sentence merges, which is important for interpreting the multilingual timing layer as high quality without pretending that the story is sentence-isomorphic across languages.

3.3 Models and sentence representations

The two model families were XLM-R-base and the encoder of NLLB-200-distilled-600M (Conneau et al., 2020; NLLB Team et al., 2024). LaBSE was used only for alignment quality control, not as a brain model. For each aligned sentence span, we ran the model with hidden states enabled, selected the tokens belonging to the current sentence, removed special tokens, mean-pooled the remaining vectors, and then L2-normalized the pooled sentence representation.

XLM-R contributed all 13 returned hidden-state layers (the embedding layer plus 12 transformer blocks). In the current archived checkpoint, NLLB brain analyses were restricted to encoder layers 4, 6, and 8 as a runtime-driven implementation deviation that is logged transparently rather than hidden post hoc. The scientific aim of the paper did not change, but this decision narrows the

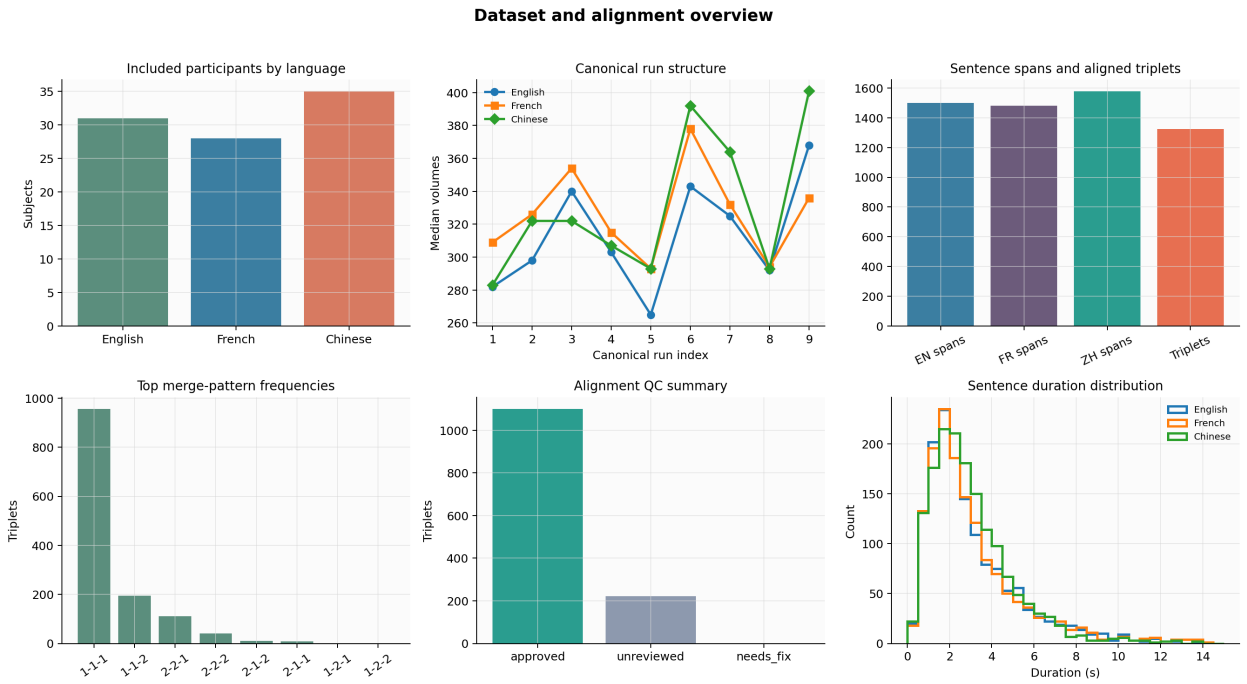

Figure 2: Dataset and alignment overview for the LPPC analysis sample. The frozen ROI bundle contains 31 English, 28 French, and 35 Chinese participants (94 total), each with nine canonical runs and TR = 2.0 s; mean story durations are 93.87, 97.90, and 99.23 minutes for English, French, and Chinese, respectively. The final tri-lingual alignment table contains 1323 triplets, with 284 rows flagged for review, 206 conflict windows repaired by the local tri-lingual reconciliation pass, and one documented unresolved outlier. Most rows are strict 1–1–1 alignments (956), but the final table also contains admissible merge patterns such as 1–1–2 (194) and 2–2–1 (112), showing that alignment quality is high without assuming perfect sentence isomorphism.

layerwise resolution of the NLLB results in the present bundle.

3.4 Shared and specific decomposition

For model m, layer l, aligned triplet s, and target language ℓ, let the normalized sentence embedding be $\tilde{h}_{s,\ell}^{(m,l)}$. We defined the leave-target-out shared component as

$$u_{s,\ell}^{(m,l)} = \frac{1}{L-1} \sum_{j \neq \ell} \tilde{h}_{s,j}^{(m,l)}, \tag{1}$$

with $L = 3$ languages. The raw residual was

$$v_{s,\ell}^{(m,l)} = \tilde{h}_{s,\ell}^{(m,l)} - u_{s,\ell}^{(m,l)}, \tag{2}$$

and the orthogonalized specific residual was

$$v_{s,\ell}^{\perp(m,l)} = v_{s,\ell}^{(m,l)} - \frac{\left(u_{s,\ell}^{(m,l)}\right)^{\top} v_{s,\ell}^{(m,l)}}{\left\| u_{s,\ell}^{(m,l)} \right\|_2^2 + \epsilon} u_{s,\ell}^{(m,l)}, \tag{3}$$

Table 2: Multilingual encoders used in the paper. LaBSE was used only for alignment quality control and was not entered into the brain-encoding analyses.

Model	Hidden size	Returned layers	Architecture	Role in paper
XLM-R-base	768	13	Multilingual encoder	Full layerwise model in the ROI analyses
NLLB-200-distilled-600M encoder	1024	13	Seq2seq encoder	Replication model; current checkpoint analyzes encoder layers 4, 6, and 8

with $\epsilon = 10^{-8}$.

The canonical feature families were `RAW` $= \tilde{h}$, `SHARED` $= u$, `SPECIFIC` $= v^{\perp}$, and `FULL` $= [u; v^{\perp}]$. `MISMATCHED_SHARED` used the correct shared vectors but shuffled triplet identities within the target canonical run before timeline placement. In the present checkpoint, the subject-level fast-paper runs used one run-local mismatch shuffle rather than the fuller default contemplated in the project specification; this is one of the logged implementation deviations carried into the archived bundle.

3.5 Design matrices, ROIs, and encoding models

All text features, including `SHARED`, were placed on the target listener's own sentence timeline rather than on the timing of the source languages. For each feature dimension we constructed a 10 Hz fine-grid boxcar, convolved it with a canonical hemodynamic response function, and sampled it at TR = 2 s. When post-story scans remained in the derivatives, the design stayed at zero after audio offset.

The nuisance design included run intercepts, run-wise linear trends, a 128-s high-pass cosine basis, RMS envelope, pitch/f_0, word rate, sentence-onset impulses, and run-1 picture regressors for English and Chinese when relevant. Motion regressors were not available in the shipped derivatives and were therefore not imputed or reconstructed; we keep that omission explicit throughout the manuscript. The acoustic-only baseline remained a separate predictive design rather than a nuisance term subtracted away inside the text-model residualization.

We used a Harvard–Oxford cortical atlas resampled with nearest-neighbor interpolation to LPPC derivative space. The primary ROI set contained 18 bilateral regions grouped into semantic, auditory/form-heavier, and control families. Semantic ROIs included bilateral posterior middle temporal gyrus, angular gyrus, temporal pole, and inferior frontal gyrus pars triangularis. Auditory/form-heavier ROIs included bilateral Heschl's gyrus, posterior superior temporal gyrus, and anterior superior temporal gyrus. Control ROIs included bilateral precentral gyrus and occipital pole. ROI targets were built by z-scoring voxel time series within run and averaging across voxels in each parcel.

Encoding used outer leave-one-run-out cross-validation over the nine canonical runs. Within each outer split, both features and targets were residualized using Frisch–Waugh–Lovell regression fit on the training runs only. Features were then standardized using training statistics, PCA was fit on training data only, and we retained the smaller of 128 principal components or the number required to explain 95% of training variance. Ridge α was selected by inner leave-one-run-out cross-validation on the eight training runs, searching 15 log-spaced values from 10^{-2} to 10^{6} and

scoring candidates by mean Fisher-z transformed held-out Pearson r, with ties broken toward the larger α. The primary subject-level summary for each condition, ROI, and layer was the mean of the nine outer-fold Fisher-z values. R^2 was treated as secondary and computed only from concatenated held-out predictions.

3.6 Confirmatory and secondary statistics

The confirmatory effect for each model, language, ROI family, and middle-late layer set was

$$\Delta z = z^{\texttt{SHARED}} - z^{\texttt{SPECIFIC}}, \tag{4}$$

with middle-late layers defined as $0.33 \leq \text{layer\_depth} \leq 0.83$. H1 tested whether the semantic-family middle-late effect was greater than zero. H2 tested whether the semantic-family shared advantage exceeded the auditory-family shared advantage. We used one-sided subject-level sign-flip permutation tests within each model-by-language slice and applied Holm correction across all 12 primary tests. The canonical paper-level aggregation reran 10,000 permutations and 10,000 bootstraps on the merged subject tables, even though the subject-level readouts were inherited from the fast-paper checkpoint.

Secondary analyses were deliberately framed more cautiously than H1. These included per-ROI contrasts, layerwise profiles, text-side multilingual geometry, geometry-to-brain coupling, and a descriptive cortex-wide localization panel. The whole-cortex panel is an ROI-projected atlas visualization rather than a voxelwise inferential encoding analysis and must be interpreted accordingly.

3.7 Transparent checkpoint note

The archived evidence bundle used for this paper inherits two substantive runtime-driven implementation deviations: subject-level `MISMATCHED_SHARED` controls were generated with one run-local shuffle, and NLLB ROI encoding was restricted to layers 4, 6, and 8. These decisions are logged in the project provenance and remain visible in the manuscript because they affect interpretive caution, especially for robustness and for the descriptive geometry-to-brain analysis. They do not change the paper's core thesis, but they do define the scope of the present checkpoint.

4 Results

4.1 Dataset integrity and alignment quality

The multilingual analysis sample is substantial but not fully symmetric across languages. The full derivative inventory contains 112 LPPC participants and 1008 run-level rows, whereas the frozen ROI bundle used for the current manuscript contains 94 participants with complete ROI targets: 31 English, 28 French, and 35 Chinese listeners. The triplet-alignment layer is similarly strong but nontrivial (Figure 2). The final alignment table contains 1323 triplets, of which 956 are strict 1–1–1 alignments; the remainder include admissible two-sentence merges such as 1–1–2 and 2–2–1. Quality control flagged 284 rows for review, repaired 206 local conflict windows, and left one documented unresolved outlier. The practical implication is that the multilingual sentence timing is usable for model–brain alignment, but the analysis does not rely on the fiction of perfect sentence isomorphism.

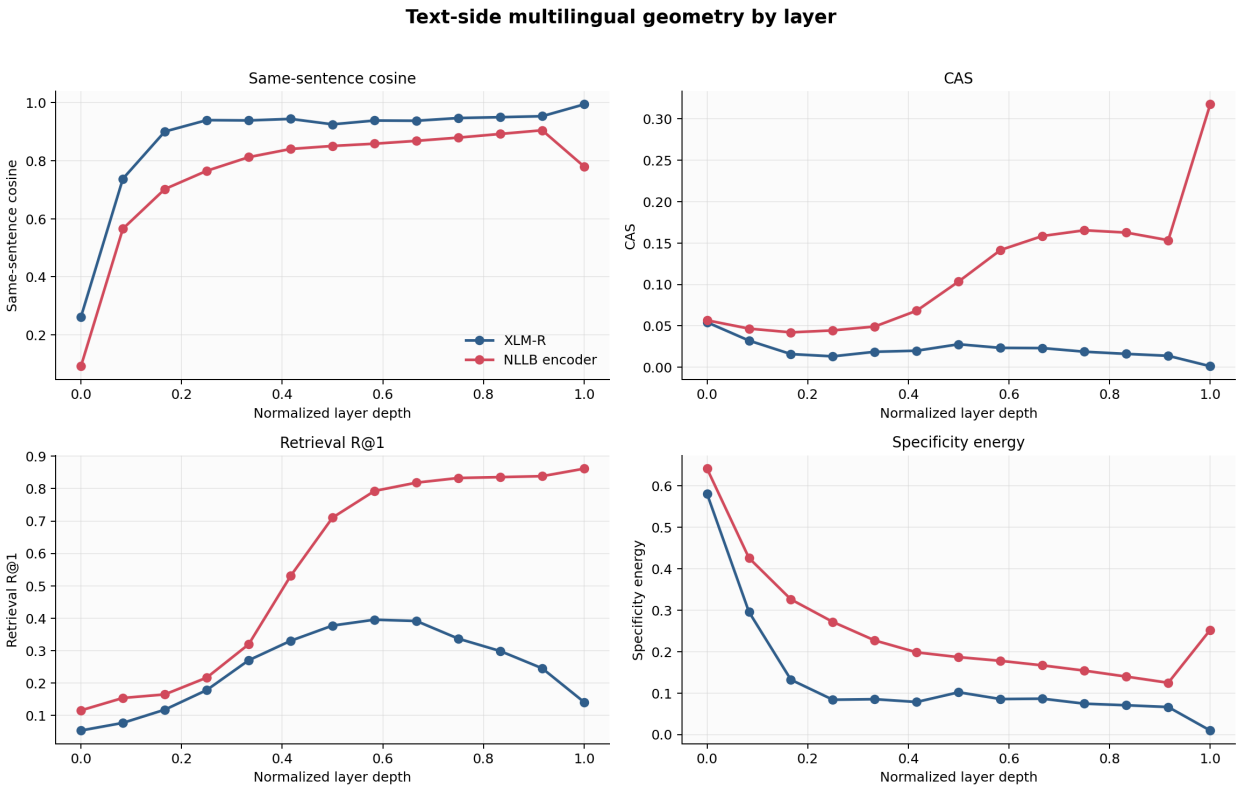

Figure 3: Layerwise multilingual geometry for XLM-R and NLLB encoder sentence representations. XLM-R reaches very high late-layer same-sentence cross-language alignment (align_mean = 0.9936 at layer 12), while retrieval peaks earlier at layer 7 ($R@1 = 0.3952$). NLLB peaks later, with CAS = 0.3174 and $R@1 = 0.8617$ at layer 12, and lowest specificity energy at layer 11 (0.1248). These metrics motivate the decomposition but are treated as descriptive rather than confirmatory.

4.2 Text-side multilingual geometry

Both model families exhibit strong cross-lingual convergence, but with different layerwise profiles (Figure 3). XLM-R becomes highly aligned in the upper stack, reaching same-sentence cross-language alignment align_mean = 0.9936 at layer 12, while its retrieval performance peaks earlier at layer 7 ($R@1 = 0.3952$). Its specificity energy declines steadily and reaches 0.00955 at layer 12, indicating that late XLM-R representations are strongly dominated by the shared axis.

NLLB shows a later and more contrastive alignment profile. Its best contrastive alignment score is 0.3174 at layer 12, retrieval reaches 0.8617 at layer 12, and alignment remains high through layers 11–12. Specificity energy bottoms out at 0.1248 at layer 11, leaving more residual language-specific energy than late XLM-R while still showing strong multilingual convergence. These text-side metrics are descriptive rather than inferential, but they establish that both model families contain the structure required for a brain-side shared-versus-specific test.

4.3 Confirmatory ROI-family results

The paper's central result is a clean positive H1 pattern (Figure 4 and table 3). In the semantic ROI family, the middle-late `SHARED`–`SPECIFIC` effect was positive in all six model-by-language tests,

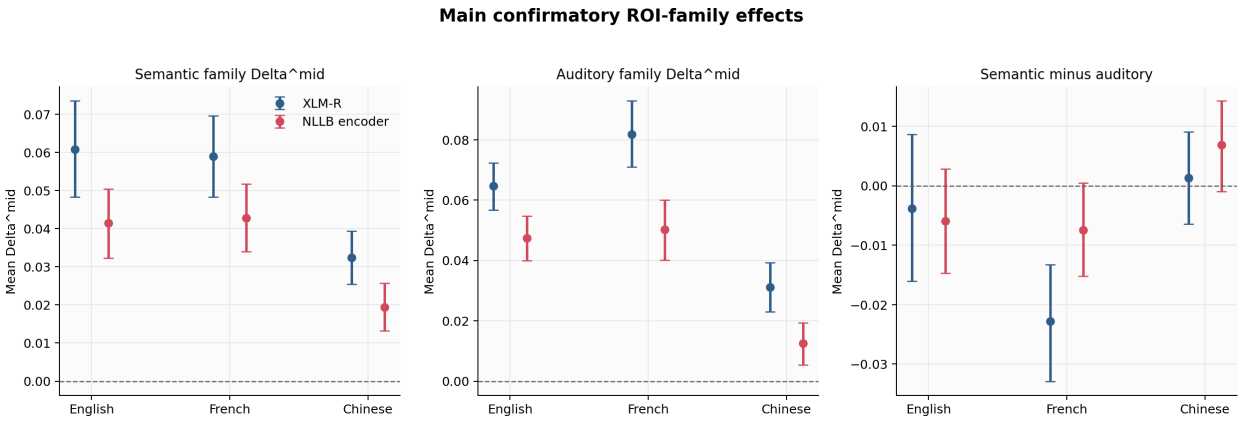

Figure 4: Main confirmatory ROI-family analyses. The semantic-family shared advantage is positive in all six model-by-language cells and remains significant after Holm correction (H1). The semantic-minus-auditory comparison (H2) is mixed, because auditory and superior temporal regions also show substantial `SHARED`–`SPECIFIC` effects. Error bars denote bootstrap confidence intervals from the archived subject-level tables.

and every one survived Holm correction. For XLM-R, the semantic-family Δz_{mid} values were 0.0609 (English), 0.0589 (French), and 0.0324 (Chinese). For NLLB they were 0.0415, 0.0429, and 0.0194. The corresponding one-sided permutation tests all yielded $p_{\text{perm}} = 0.0001$, and all six Holm-adjusted values were 0.0012.

By contrast, H2 was mixed rather than confirmatory. XLM-R yielded semantic-minus-auditory differences of −0.0038 (English), −0.0228 (French), and 0.0013 (Chinese). NLLB yielded −0.0059, −0.0074, and 0.0069, with the Chinese row nominally one-sided positive before correction but not Holm-significant. The paper's main positive claim is therefore the semantic-family `SHARED` > `SPECIFIC` effect itself, not a strong semantic-versus-auditory dissociation.

4.4 Per-ROI condition comparisons

The ROI-family result is not driven by a small subset of parcels. Every semantic ROI row in the full 48-row semantic block had a positive `SHARED`–`SPECIFIC` effect and all 48 survived FDR correction. Examples include XLM-R English `R_AG` ($\Delta z = 0.0786$, $q_{\text{FDR}} = 0.000138$), XLM-R French `R_IFGtri` (0.0759, $q = 0.000138$), NLLB English `R_AG` (0.0495, $q = 0.000138$), and NLLB French `R_IFGtri` (0.0698, $q = 0.000164$).

Auditory and superior temporal ROIs were also strongly positive. Across the 36 auditory rows, 35 survived FDR correction, with the largest effects in XLM-R French `R_pSTG` (0.1096, $q = 0.000138$) and NLLB French `R_pSTG` (0.0719, $q = 0.000164$). This is the immediate reason H2 stays mixed: auditory cortex is not a near-zero fallback for the shared component.

Correct-content controls were directionally sane. `SHARED` exceeded `MISMATCHED_SHARED` in every semantic ROI row (48/48) and every auditory row (36/36). Control-family effects were weaker, as they should be: 23 of 24 control rows were positive and 11 of 24 survived FDR correction, with one near-zero negative row in XLM-R Chinese `L_OccipitalPole`. Quality-control reports also showed that `RAW` outperformed the acoustic-only baseline somewhere in the language-sensitive ROI set for both model families.

Table 3: Confirmatory ROI-family tests. H1 asks whether the semantic-family middle-late shared advantage is greater than zero. H2 asks whether the semantic-family shared advantage exceeds the auditory-family shared advantage.

Model	Lang.	Hyp.	Δz_{mid}	95% CI	p_{perm}	Holm p
XLM-R	EN	H1	0.0609	[0.0483, 0.0738]	0.0001	0.0012
XLM-R	EN	H2	-0.0038	[-0.0161, 0.0087]	0.7234	1.0000
XLM-R	FR	H1	0.0589	[0.0484, 0.0697]	0.0001	0.0012
XLM-R	FR	H2	-0.0228	[-0.0329, -0.0133]	1.0000	1.0000
XLM-R	ZH	H1	0.0324	[0.0255, 0.0393]	0.0001	0.0012
XLM-R	ZH	H2	0.0013	[-0.0065, 0.0091]	0.3634	1.0000
NLLB encoder	EN	H1	0.0415	[0.0323, 0.0505]	0.0001	0.0012
NLLB encoder	EN	H2	-0.0059	[-0.0147, 0.0028]	0.9008	1.0000
NLLB encoder	FR	H1	0.0429	[0.0340, 0.0517]	0.0001	0.0012
NLLB encoder	FR	H2	-0.0074	[-0.0152, 0.0005]	0.9586	1.0000
NLLB encoder	ZH	H1	0.0194	[0.0132, 0.0255]	0.0001	0.0012
NLLB encoder	ZH	H2	0.0069	[-0.0009, 0.0143]	0.0426	0.2556

All six H1 tests are positive and survive Holm correction. No H2 comparison survives Holm correction.

Table 4: Selected representative ROI contrasts derived from the full ROI condition statistics. The table highlights strong semantic examples, strong auditory examples that help explain the mixed H2 result, and weaker control-family cases.

Model	Lang.	ROI	Fam.	SHARED	SPECIFIC	MISMATCHED_SHARED	SHARED–SPECIFIC	q_{FDR}
XLM-R	EN	R_AG	Sem.	0.1244	0.0458	-0.0063	0.0786	0.0001
XLM-R	FR	R_IFGtri	Sem.	0.1071	0.0313	0.0051	0.0759	0.0001
XLM-R	FR	R_pSTG	Aud.	0.1654	0.0558	0.0755	0.1096	0.0001
XLM-R	ZH	L_OccipitalPole	Ctrl.	0.0136	0.0140	-0.0010	-0.0004	0.5148
NLLB encoder	EN	R_AG	Sem.	0.1204	0.0709	-0.0036	0.0495	0.0001
NLLB encoder	FR	R_IFGtri	Sem.	0.1028	0.0330	0.0083	0.0698	0.0002
NLLB encoder	FR	R_pSTG	Aud.	0.1616	0.0897	0.0684	0.0719	0.0002
NLLB encoder	ZH	R_Precentral	Ctrl.	0.0221	0.0193	0.0081	0.0028	0.3510

4.5 Layerwise profiles

Figure 6 shows that the brain-side shared advantage is also layer-sensitive. In the semantic family, XLM-R peaks at layer 7 for English and French (0.0681 and 0.0672) and later in Chinese at layer 10 (0.0395). NLLB semantic-family peaks are more consistent, with all three languages maximizing at layer 8 (0.0551, 0.0518, and 0.0330).

Auditory-family peaks are often later for XLM-R than for the semantic family, especially in English and French, where auditory maxima occur at layers 10 and 9. In NLLB, auditory peaks appear at layer 6 for English and French and at layer 8 for Chinese. Representative ROI examples mirror the family-level summaries: XLM-R English R_pSTG peaks at layer 10 (0.0936), whereas XLM-R English R_AG peaks earlier at layer 9 (0.0868), and NLLB semantic ROIs such as L_pMTG and R_AG peak at layer 8. These profiles reveal structure in how the shared advantage is distributed over model depth, but they do not rescue H2; if anything, they emphasize that auditory and superior temporal regions carry substantial shared-over-specific signal in middle and late layers.

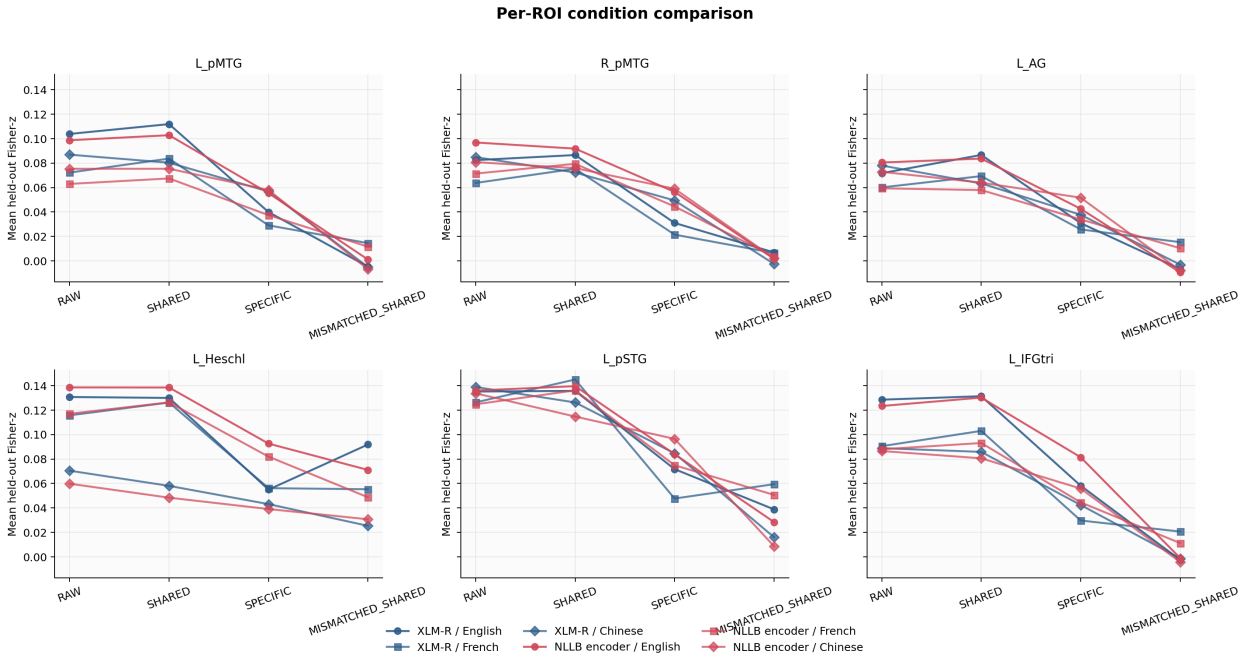

Figure 5: Representative ROI condition comparisons. Semantic examples such as angular gyrus and inferior frontal gyrus pars triangularis show a clear `SHARED` > `SPECIFIC` pattern, but auditory examples, especially posterior superior temporal gyrus, also display sizeable positive `SHARED`–`SPECIFIC` effects. This is why H2 remains mixed even though H1 replicates cleanly.

4.6 Geometry-to-brain coupling

The geometry-to-brain coupling analysis was computed, but it was not a headline success (Figure 7). XLM-R showed weak and non-significant correlations between layerwise contrastive alignment and semantic-family shared-brain advantage in all three languages ($\rho = -0.099$, 0.027, and -0.192; all $p_{\text{perm}} \geq 0.52$). NLLB reached $\rho = 1.0$ in each language, but each estimate was based on only three sampled layers and yielded the same non-significant permutation result ($p_{\text{perm}} = 0.4286$). We therefore treat geometry-to-brain coupling as an interpretable secondary analysis that remains suggestive rather than strongly supported in the present bundle.

4.7 Descriptive cortex-wide localization

Figure 8 provides cortex-wide context for the ROI findings. In the present bundle the panel was rendered as an ROI-projected atlas visualization rather than a voxelwise encoding map. The representative layers were XLM-R layer 7 and NLLB layer 8, chosen by semantic-family shared advantage. The figure is useful for showing where the ROI-level effects sit in cortical space, but it must not be interpreted as a voxelwise inferential whole-brain result.

4.8 Robustness analyses

The robustness suite refines the interpretation rather than overturning it (Figure 9 and table 5). Four representative-layer perturbations preserved the sign of the semantic-family shared advan-

Table 5: Robustness of the representative-layer semantic `SHARED`–`SPECIFIC` effect. The robustness suite is supportive but bounded rather than uniformly clean.

Condition	All 6 cells positive?	Summary
Last-token pooling	No	5/6 model-language cells stayed positive; failures: XLM-R/ZH (-0.001). Reference base was 6/6 positive.
4-lag FIR	Yes	6/6 model-language cells stayed positive. Reference base was 6/6 positive.
No acoustic nuisance	Yes	6/6 model-language cells stayed positive. Reference base was 6/6 positive.
No pitch nuisance	Yes	6/6 model-language cells stayed positive. Reference base was 6/6 positive.
Previous 2-sentence context	Yes	6/6 model-language cells stayed positive. Reference base was 6/6 positive.
Voxelwise within-ROI mean-z	No	4/6 model-language cells stayed positive; failures: XLM-R/FR (-0.029), NLLB encoder/FR (-0.025). Reference base was 6/6 positive.

tage in all 6 model-language cells: `fir_4lag`, `no_acoustic_nuisance`, `no_pitch_nuisance`, and `previous_2_sentence_context`. Last-token pooling was only slightly less stable, preserving the sign in 5 of 6 cells, with a single near-zero miss in XLM-R Chinese (-0.001).

The most demanding variant was the voxelwise-within-ROI mean-z target, which preserved the sign in 4 of 6 cells and produced negative French rows in both model families (XLM-R -0.029, NLLB encoder -0.025). We therefore interpret the robustness block as supportive but not uniform. The main semantic-family `SHARED` > `SPECIFIC` result survives several changes to temporal basis, nuisance composition, and context definition, but it is less stable when the target is moved from ROI means to within-ROI voxelwise summaries, especially in the French cohorts.

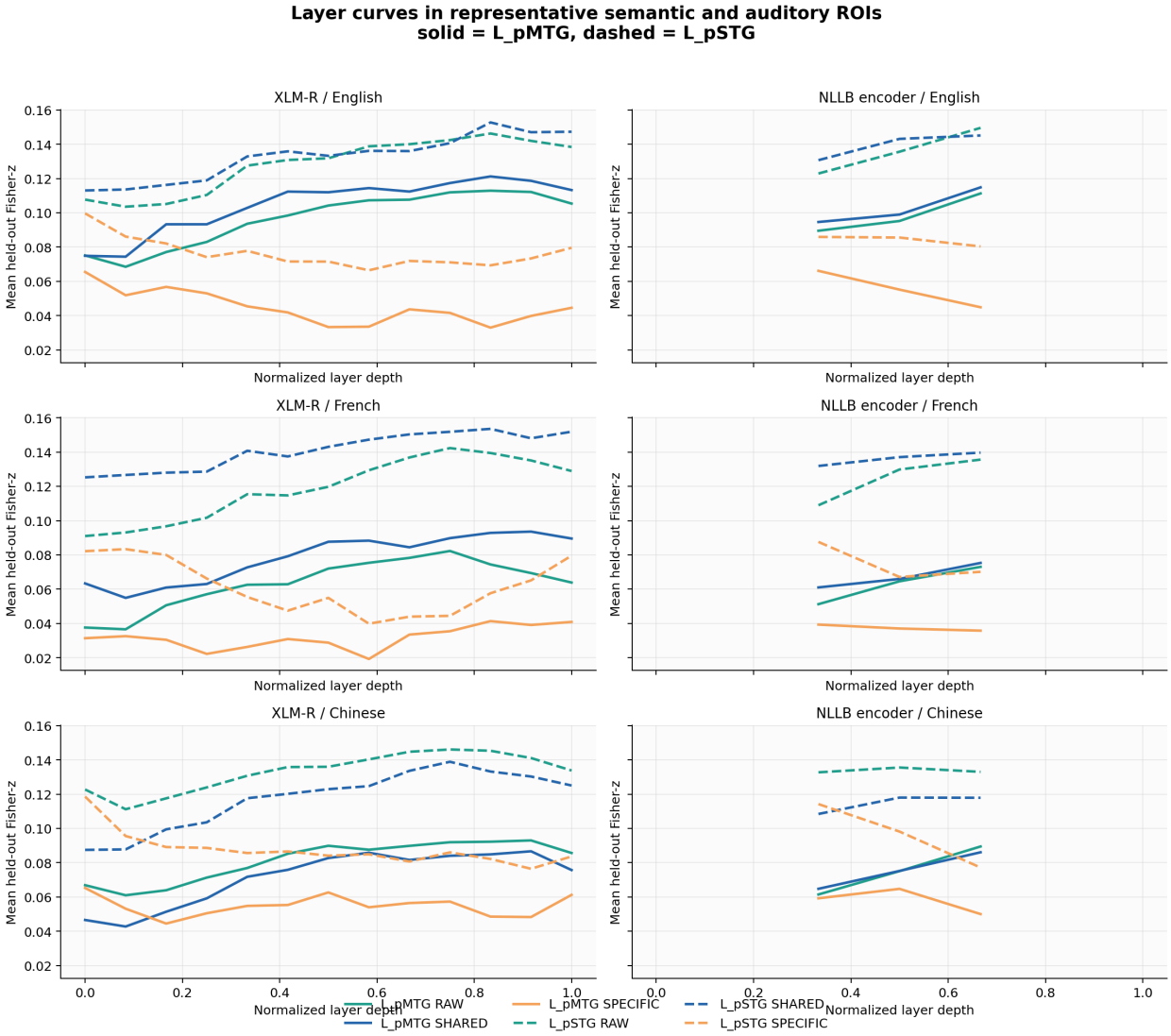

Figure 6: Layerwise profiles for representative semantic and auditory ROIs. XLM-R semantic-family effects peak at layer 7 in English and French and later in Chinese; NLLB semantic-family peaks are concentrated at layer 8. Auditory-family peaks often occur comparably late or later, especially in superior temporal cortex, which helps explain the mixed H2 results.

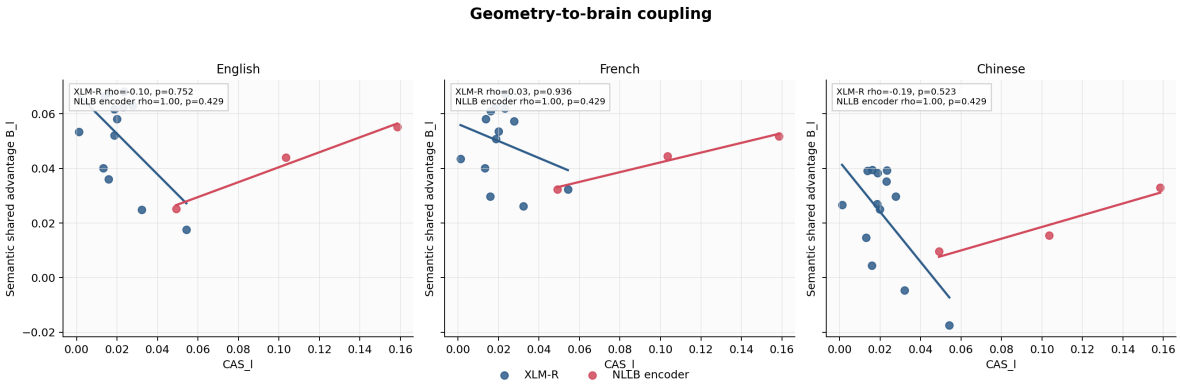

Figure 7: Geometry-to-brain coupling as a descriptive secondary analysis. XLM-R shows weak, non-significant correlations between contrastive alignment score and semantic-family shared advantage. NLLB attains monotonic three-point trends because only three encoder layers were evaluated in the current checkpoint, and the corresponding permutation tests remain non-significant.

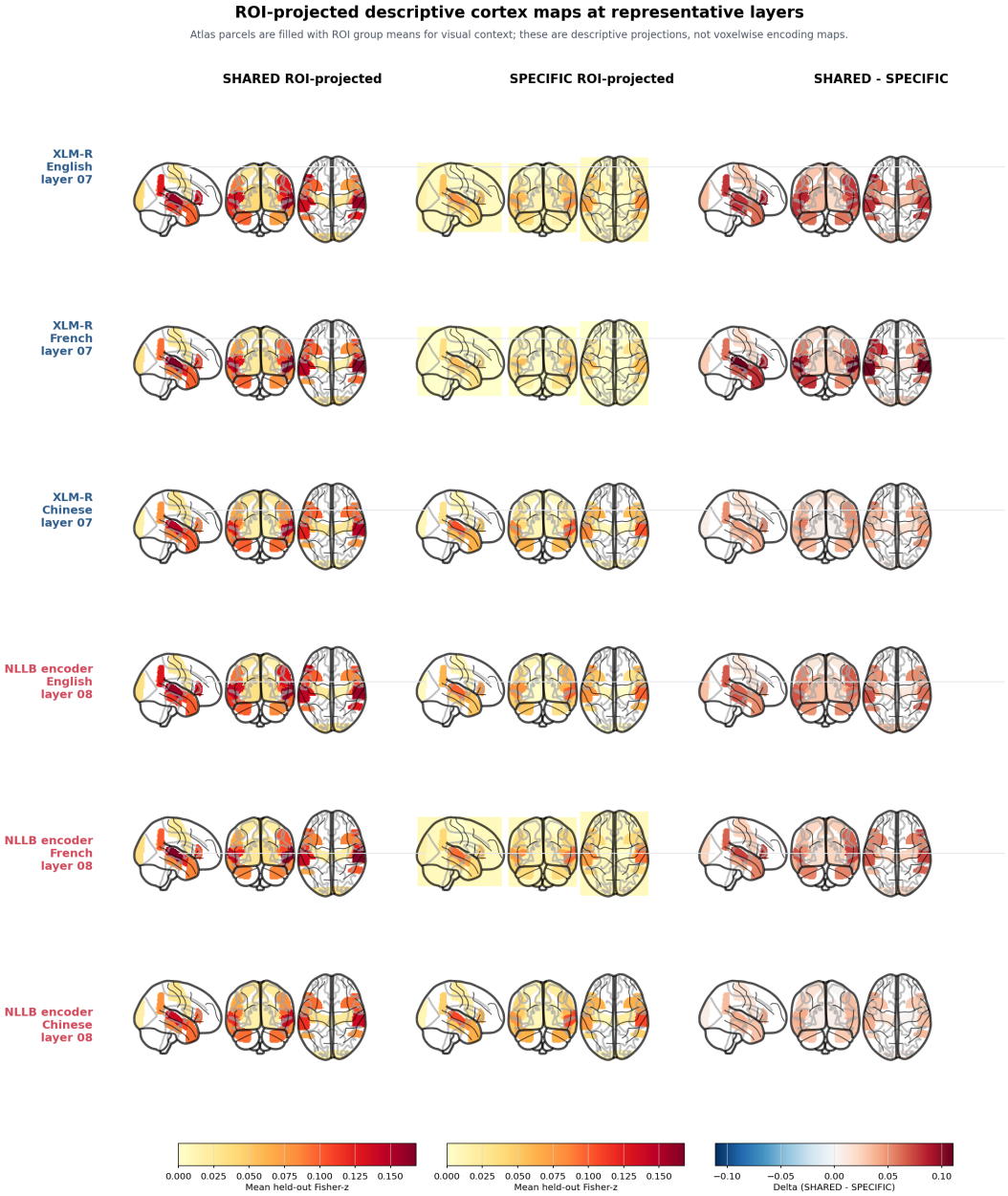

Figure 8: Descriptive ROI-projected cortex panel at the representative semantic layers (XLM-R layer 7 and NLLB layer 8). This figure offers spatial context for the ROI results but is not a voxelwise inferential analysis. The bundle's provenance files explicitly mark it as descriptive only.

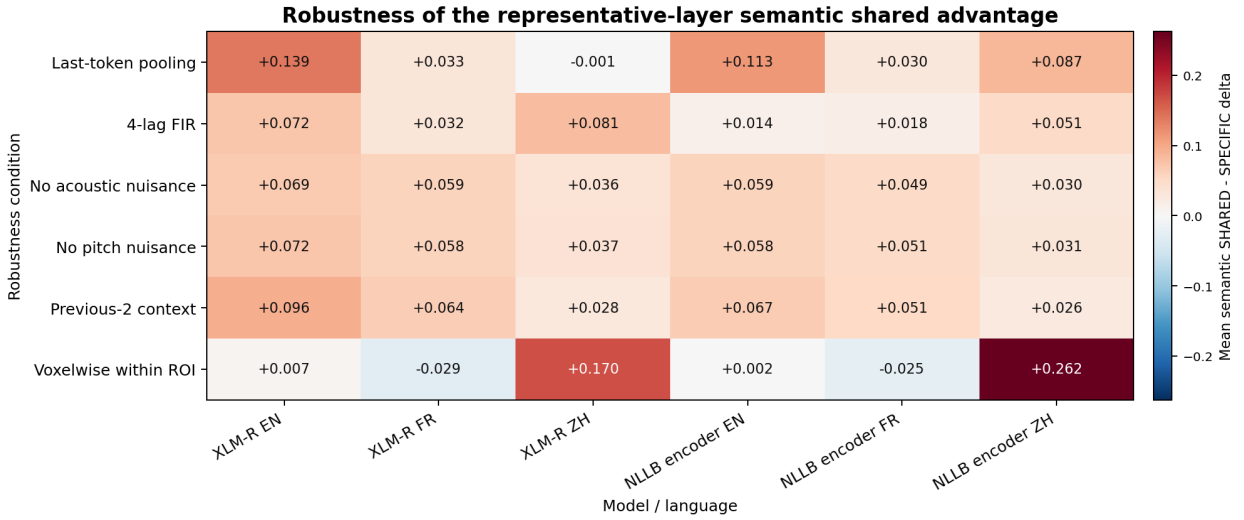

Figure 9: Robustness of the representative-layer semantic `SHARED–SPECIFIC` effect. The core sign is preserved for 6/6 cells under 4-lag FIR, no-acoustic-nuisance, no-pitch-nuisance, and previous-2-sentence context variants. Last-token pooling yields one near-zero miss in XLM-R Chinese, and the voxelwise-within-ROI target produces negative French rows in both model families, making the robustness story supportive but bounded.

5 Discussion

The main result of the paper is clear. Across two multilingual model families and three language cohorts, higher-level semantic ROIs were better predicted by a leave-target-out shared representation than by an orthogonalized target-language-specific residual. That result survived Holm correction in all six confirmatory H1 tests and therefore supports the central mechanistic claim of the project.

At the same time, the paper is narrower than the strongest version of the original story. We expected the shared-over-specific advantage to be clearly larger in the semantic family than in the auditory/form-heavier family, but that prediction remained mixed in both model families. The per-ROI breakdown shows why: superior temporal and auditory-adjacent regions, especially posterior STG and anterior STG, also showed strong shared-over-specific effects. The right interpretation is therefore not that the brain cleanly separates a shared semantic code from a specific auditory code, but that the distribution of shared and specific sensitivity varies by region and is not exhausted by a simple semantic-versus-auditory contrast.

This still makes multilingual LLMs useful explanatory probes. The decomposition into shared and specific components is doing scientific work that a raw embedding alone cannot do: it lets us ask which part of a multilingual representation better explains a target listener's responses. The fact that `SHARED` exceeded `MISMATCHED_SHARED` broadly in semantic and auditory ROI rows reinforces that the effect depends on correct content correspondence rather than on the mere existence of a smooth multilingual feature space.

The current geometry-to-brain coupling results are more modest. Text-side multilingual convergence was measurable, especially in late NLLB layers, but the coupling between convergence and semantic-family brain advantage was weak or statistically unstable. That does not undermine the core shared-versus-specific finding, but it argues against overselling the geometry analysis as evidence that model-space alignment and brain-space preference are the same phenomenon.

Overall, the present analysis supports a robust multilingual shared-versus-specific effect in higher-level language cortex, while also showing that specific residual sensitivity cannot be reduced to a single auditory fallback story. This is exactly the kind of nuanced outcome that justifies a mechanistic paper rather than a benchmark paper.

6 Limitations and Transparency

Several limitations should remain explicit.

First, the analysis is sentence-based. This keeps multilingual alignment and design construction tractable, but it approximates discourse with sentence spans rather than modeling long-context comprehension directly.

Second, the ROI analysis uses an anatomical atlas rather than subject-specific functional ROIs. That choice improves reproducibility but introduces mismatch between atlas anatomy and each participant's functional organization.

Third, the atlas is resampled from a standard MNI152-distributed product into LPPC `MNIColin27` derivative space. This template mismatch is manageable for an ROI-first analysis, but it is still a source of anatomical imprecision.

Fourth, the English, French, and Chinese cohorts come from different acquisition settings. We therefore do not treat cross-language magnitude differences as scanner-free truths, and we keep pooled cross-language interpretation descriptive.

Fifth, motion regressors were not available in the shipped derivative bundle used here. We therefore omitted them rather than introducing a new reprocessing pipeline, and that omission should remain visible in the manuscript.

Sixth, the analysis is correlative. Even a replicated `SHARED` > `SPECIFIC` effect does not show that the brain implements the same representations as the models or that the factorization is causally privileged.

Seventh, the current cortex-wide panel is descriptive. Figure 8 is an ROI-projected atlas visualization, not a voxelwise confirmatory map.

Eighth, the present manuscript checkpoint still inherits a fast-paper run definition: one run-local mismatched shuffle at the subject level and sparse NLLB layer coverage (4, 6, 8). The confirmatory paper-level statistics were recomputed rigorously from the merged subject tables, but the current checkpoint should still be refreshed if a denser final run definition is chosen.

Ninth, the completed robustness suite is supportive but not perfectly clean. The voxelwise-within-ROI target produces negative French rows in both model families, so robustness should be treated as a bounded caveat on generality rather than as a uniformly positive add-on to the headline claim.

7 Conclusion

In a multilingual naturalistic fMRI setting, a leave-target-out shared multilingual representation predicted semantic-language ROI responses better than a target-language-specific residual across both XLM-R and NLLB and across English, French, and Chinese listeners. The stronger semantic-over-auditory dissociation remained mixed, which narrows the claim but does not erase it. The most defensible conclusion from the present evidence bundle is that multilingual LLMs can act as factorized probes of naturalistic language processing, with higher-level language cortex showing reliable sensitivity to cross-lingually shared structure and a more distributed regional pattern than a simple semantic-versus-auditory split would suggest.

Data and Code Availability

The core dataset is LPPC on OpenNeuro (`ds003643`) (Li et al., 2022). All analysis code, generated tables, figures, logs, and manuscript assets for the archived checkpoint are contained in the accompanying source package. The canonical evidence bundle for the manuscript is anchored by the dataset summary, model summary, confirmatory statistics, ROI condition statistics, robustness summary, provenance logs, and the canonical figures reused here.

References

Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux. Machine learning for

neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8:14, 2014. doi: 10.3389/fninf.2014.00014.

Charlotte Caucheteux and Jean-Rémi King. Brains and algorithms partially converge in natural language processing. *Communications Biology*, 5:134, 2022. doi: 10.1038/s42003-022-03036-1.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, 2020. doi: 10.18653/v1/2020.acl-main.747.

Jixing Li, Shohini Bhattasali, Shumin Zhang, Brent Franzluebbers, Wen-Ming Luh, R. Nathan Spreng, Jonathan R. Brennan, Yiming Yang, Christophe Pallier, and John Hale. Le petit prince multilingual naturalistic fmri corpus. *Scientific Data*, 9(1):436, 2022. doi: 10.1038/s41597-022-01551-0.

Samuel A. Nastase, Yibei Liu, Hannah Hillman, Kenneth A. Norman, and Uri Hasson. The "narratives" fmri dataset for evaluating models of naturalistic language comprehension. *Scientific Data*, 8:250, 2021. doi: 10.1038/s41597-021-01033-3.

NLLB Team, Marta R. Costa-jussà, Anwar Hasan, et al. Scaling neural machine translation to 200 languages. *Nature*, 630:841–846, 2024. doi: 10.1038/s41586-024-07335-x.

Martin Schrimpf, Idan Asher Blank, Greta Tuckute, et al. The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, 118(45):e2105646118, 2021. doi: 10.1073/pnas.2105646118.

Mariya Toneva and Leila Wehbe. Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). In *Advances in Neural Information Processing Systems 32*, pages 14928–14938, 2019.

A Supplementary ROI statistics

Table 6 provides the full 108-row ROI condition statistics table underlying the selected rows in Table 4 and the condition comparisons in Figure 5.

Table 6. Full ROI condition statistics underlying Figure 5 and Table 4. Values are middle-late-layer summaries from the archived fast-paper checkpoint.

Model	Lang.	ROI	Fam.	SHARED	SPECIFIC	MISMATCHED_SHARED	SHARED–SPECIFIC	q_{FDR}
NLLB encoder	EN	L_Heschl	Aud.	0.1385	0.0926	0.0711	0.0459	0.0001
NLLB encoder	EN	L_aSTG	Aud.	0.1071	0.0579	0.0270	0.0492	0.0001
NLLB encoder	EN	L_pSTG	Aud.	0.1397	0.0840	0.0284	0.0557	0.0001
NLLB encoder	EN	R_Heschl	Aud.	0.1093	0.0726	0.0476	0.0367	0.0001
NLLB encoder	EN	R_aSTG	Aud.	0.1074	0.0674	0.0407	0.0400	0.0001
NLLB encoder	EN	R_pSTG	Aud.	0.1637	0.1068	0.0645	0.0569	0.0001
NLLB encoder	EN	L_OccipitalPole	Ctrl.	0.0333	0.0002	-0.0039	0.0331	0.0004
NLLB encoder	EN	L_Precentral	Ctrl.	0.0218	0.0112	-0.0053	0.0107	0.0731
NLLB encoder	EN	R_OccipitalPole	Ctrl.	0.0282	0.0090	-0.0015	0.0193	0.0017
NLLB encoder	EN	R_Precentral	Ctrl.	0.0123	0.0046	-0.0130	0.0077	0.1534
NLLB encoder	EN	L_AG	Sem.	0.0837	0.0426	-0.0094	0.0411	0.0001
NLLB encoder	EN	L_IFGtri	Sem.	0.1303	0.0815	-0.0010	0.0488	0.0001
NLLB encoder	EN	L_TemporalPole	Sem.	0.0903	0.0564	0.0058	0.0339	0.0001
NLLB encoder	EN	L_pMTG	Sem.	0.1028	0.0554	0.0011	0.0474	0.0001
NLLB encoder	EN	R_AG	Sem.	0.1204	0.0709	-0.0036	0.0495	0.0001
NLLB encoder	EN	R_IFGtri	Sem.	0.1034	0.0680	-0.0035	0.0355	0.0001
NLLB encoder	EN	R_TemporalPole	Sem.	0.0784	0.0383	-0.0024	0.0401	0.0001
NLLB encoder	EN	R_pMTG	Sem.	0.0918	0.0564	0.0016	0.0354	0.0003
NLLB encoder	FR	L_Heschl	Aud.	0.1264	0.0820	0.0487	0.0444	0.0002
NLLB encoder	FR	L_aSTG	Aud.	0.1107	0.0664	0.0293	0.0442	0.0002
NLLB encoder	FR	L_pSTG	Aud.	0.1362	0.0749	0.0506	0.0613	0.0002
NLLB encoder	FR	R_Heschl	Aud.	0.0968	0.0645	0.0375	0.0323	0.0002
NLLB encoder	FR	R_aSTG	Aud.	0.1159	0.0683	0.0455	0.0476	0.0002
NLLB encoder	FR	R_pSTG	Aud.	0.1616	0.0897	0.0684	0.0719	0.0002
NLLB encoder	FR	L_OccipitalPole	Ctrl.	0.0247	0.0140	-0.0035	0.0107	0.1371
NLLB encoder	FR	L_Precentral	Ctrl.	0.0203	0.0114	0.0026	0.0089	0.1884
NLLB encoder	FR	R_OccipitalPole	Ctrl.	0.0325	0.0134	-0.0022	0.0191	0.0258
NLLB encoder	FR	R_Precentral	Ctrl.	0.0159	0.0032	-0.0023	0.0127	0.0618
NLLB encoder	FR	L_AG	Sem.	0.0579	0.0339	0.0103	0.0240	0.0037
NLLB encoder	FR	L_IFGtri	Sem.	0.0931	0.0446	0.0112	0.0486	0.0002
NLLB encoder	FR	L_TemporalPole	Sem.	0.0788	0.0329	0.0089	0.0460	0.0002
NLLB encoder	FR	L_pMTG	Sem.	0.0674	0.0374	0.0116	0.0300	0.0004
NLLB encoder	FR	R_AG	Sem.	0.0992	0.0497	0.0148	0.0495	0.0002
NLLB encoder	FR	R_IFGtri	Sem.	0.1028	0.0330	0.0083	0.0698	0.0002
NLLB encoder	FR	R_TemporalPole	Sem.	0.0798	0.0396	0.0144	0.0401	0.0003
NLLB encoder	FR	R_pMTG	Sem.	0.0794	0.0445	0.0034	0.0349	0.0002
NLLB encoder	ZH	L_Heschl	Aud.	0.0484	0.0391	0.0307	0.0093	0.0428
NLLB encoder	ZH	L_aSTG	Aud.	0.0771	0.0593	0.0145	0.0178	0.0081
NLLB encoder	ZH	L_pSTG	Aud.	0.1148	0.0965	0.0087	0.0183	0.0061
NLLB encoder	ZH	R_Heschl	Aud.	0.0358	0.0311	0.0245	0.0047	0.2585
NLLB encoder	ZH	R_aSTG	Aud.	0.1038	0.0917	0.0257	0.0121	0.0235

Continued on next page

Table 6 (continued)

Model	Lang.	ROI	Fam.	SHARED	SPECIFIC	MISMATCHED_SHARED	SHARED–SPECIFIC	q_{FDR}
NLLB encoder	ZH	R_pSTG	Aud.	0.1369	0.1242	0.0311	0.0128	0.0149
NLLB encoder	ZH	L_OccipitalPole	Ctrl.	0.0147	0.0114	-0.0055	0.0033	0.3137
NLLB encoder	ZH	L_Precentral	Ctrl.	0.0271	0.0131	0.0070	0.0140	0.0495
NLLB encoder	ZH	R_OccipitalPole	Ctrl.	0.0137	0.0085	-0.0025	0.0052	0.2529
NLLB encoder	ZH	R_Precentral	Ctrl.	0.0221	0.0193	0.0081	0.0028	0.3510
NLLB encoder	ZH	L_AG	Sem.	0.0640	0.0517	-0.0079	0.0123	0.0108
NLLB encoder	ZH	L_IFGtri	Sem.	0.0806	0.0559	-0.0042	0.0248	0.0009
NLLB encoder	ZH	L_TemporalPole	Sem.	0.0953	0.0742	-0.0078	0.0211	0.0081
NLLB encoder	ZH	L_pMTG	Sem.	0.0754	0.0580	-0.0068	0.0174	0.0009
NLLB encoder	ZH	R_AG	Sem.	0.0769	0.0482	0.0010	0.0287	0.0012
NLLB encoder	ZH	R_IFGtri	Sem.	0.0895	0.0727	0.0057	0.0168	0.0075
NLLB encoder	ZH	R_TemporalPole	Sem.	0.0952	0.0783	0.0006	0.0170	0.0113
NLLB encoder	ZH	R_pMTG	Sem.	0.0758	0.0590	0.0022	0.0168	0.0061
XLM-R	EN	L_Heschl	Aud.	0.1300	0.0550	0.0919	0.0750	0.0001
XLM-R	EN	L_aSTG	Aud.	0.0967	0.0426	0.0385	0.0541	0.0001
XLM-R	EN	L_pSTG	Aud.	0.1358	0.0717	0.0388	0.0641	0.0001
XLM-R	EN	R_Heschl	Aud.	0.1002	0.0388	0.0698	0.0614	0.0001
XLM-R	EN	R_aSTG	Aud.	0.1039	0.0473	0.0528	0.0566	0.0001
XLM-R	EN	R_pSTG	Aud.	0.1584	0.0817	0.0786	0.0767	0.0001
XLM-R	EN	L_OccipitalPole	Ctrl.	0.0426	0.0102	0.0071	0.0324	0.0020
XLM-R	EN	L_Precentral	Ctrl.	0.0219	-0.0001	-0.0016	0.0220	0.0025
XLM-R	EN	R_OccipitalPole	Ctrl.	0.0360	0.0122	0.0030	0.0238	0.0030
XLM-R	EN	R_Precentral	Ctrl.	0.0209	-0.0014	-0.0108	0.0223	0.0020
XLM-R	EN	L_AG	Sem.	0.0867	0.0309	-0.0067	0.0558	0.0001
XLM-R	EN	L_IFGtri	Sem.	0.1314	0.0582	-0.0025	0.0732	0.0001
XLM-R	EN	L_TemporalPole	Sem.	0.0940	0.0431	0.0071	0.0509	0.0001
XLM-R	EN	L_pMTG	Sem.	0.1119	0.0399	-0.0048	0.0720	0.0001
XLM-R	EN	R_AG	Sem.	0.1244	0.0458	-0.0063	0.0786	0.0001
XLM-R	EN	R_IFGtri	Sem.	0.0964	0.0493	-0.0031	0.0471	0.0003
XLM-R	EN	R_TemporalPole	Sem.	0.0732	0.0195	0.0074	0.0537	0.0001
XLM-R	EN	R_pMTG	Sem.	0.0866	0.0310	0.0069	0.0556	0.0001
XLM-R	FR	L_Heschl	Aud.	0.1263	0.0562	0.0553	0.0701	0.0001
XLM-R	FR	L_aSTG	Aud.	0.1128	0.0447	0.0414	0.0681	0.0001
XLM-R	FR	L_pSTG	Aud.	0.1451	0.0477	0.0594	0.0974	0.0001
XLM-R	FR	R_Heschl	Aud.	0.1037	0.0397	0.0396	0.0640	0.0001
XLM-R	FR	R_aSTG	Aud.	0.1191	0.0379	0.0523	0.0812	0.0001
XLM-R	FR	R_pSTG	Aud.	0.1654	0.0558	0.0755	0.1096	0.0001
XLM-R	FR	L_OccipitalPole	Ctrl.	0.0210	0.0107	-0.0018	0.0103	0.1366
XLM-R	FR	L_Precentral	Ctrl.	0.0211	0.0063	0.0044	0.0148	0.0300
XLM-R	FR	R_OccipitalPole	Ctrl.	0.0298	0.0103	0.0077	0.0195	0.0250
XLM-R	FR	R_Precentral	Ctrl.	0.0102	0.0019	0.0045	0.0083	0.1504
XLM-R	FR	L_AG	Sem.	0.0693	0.0257	0.0153	0.0436	0.0003
XLM-R	FR	L_IFGtri	Sem.	0.1031	0.0297	0.0206	0.0733	0.0001

Continued on next page

Table 6 (continued)

Model	Lang.	ROI	Fam.	SHARED	SPECIFIC	MISMATCHED_SHARED	SHARED–SPECIFIC	q_{FDR}
XLM-R	FR	L_TemporalPole	Sem.	0.0902	0.0238	0.0107	0.0665	0.0001
XLM-R	FR	L_pMTG	Sem.	0.0837	0.0290	0.0145	0.0546	0.0001
XLM-R	FR	R_AG	Sem.	0.0931	0.0409	0.0154	0.0522	0.0001
XLM-R	FR	R_IFGtri	Sem.	0.1071	0.0313	0.0051	0.0759	0.0001
XLM-R	FR	R_TemporalPole	Sem.	0.0797	0.0283	0.0128	0.0514	0.0001
XLM-R	FR	R_pMTG	Sem.	0.0754	0.0215	0.0064	0.0539	0.0001
XLM-R	ZH	L_Heschl	Aud.	0.0581	0.0431	0.0254	0.0149	0.0021
XLM-R	ZH	L_aSTG	Aud.	0.0823	0.0510	0.0191	0.0312	0.0002
XLM-R	ZH	L_pSTG	Aud.	0.1263	0.0847	0.0161	0.0417	0.0002
XLM-R	ZH	R_Heschl	Aud.	0.0442	0.0307	0.0209	0.0135	0.0149
XLM-R	ZH	R_aSTG	Aud.	0.1084	0.0676	0.0237	0.0408	0.0002
XLM-R	ZH	R_pSTG	Aud.	0.1467	0.1022	0.0322	0.0445	0.0002
XLM-R	ZH	L_OccipitalPole	Ctrl.	0.0136	0.0140	-0.0010	-0.0004	0.5148
XLM-R	ZH	L_Precentral	Ctrl.	0.0224	0.0057	0.0048	0.0166	0.0094
XLM-R	ZH	R_OccipitalPole	Ctrl.	0.0135	0.0125	0.0014	0.0011	0.4752
XLM-R	ZH	R_Precentral	Ctrl.	0.0204	0.0105	0.0092	0.0099	0.0848
XLM-R	ZH	L_AG	Sem.	0.0634	0.0377	-0.0031	0.0257	0.0002
XLM-R	ZH	L_IFGtri	Sem.	0.0858	0.0424	-0.0013	0.0434	0.0002
XLM-R	ZH	L_TemporalPole	Sem.	0.0990	0.0662	-0.0050	0.0328	0.0002
XLM-R	ZH	L_pMTG	Sem.	0.0803	0.0568	-0.0052	0.0235	0.0003
XLM-R	ZH	R_AG	Sem.	0.0773	0.0316	0.0014	0.0457	0.0002
XLM-R	ZH	R_IFGtri	Sem.	0.0890	0.0560	0.0009	0.0330	0.0002
XLM-R	ZH	R_TemporalPole	Sem.	0.0926	0.0601	-0.0036	0.0325	0.0002
XLM-R	ZH	R_pMTG	Sem.	0.0723	0.0494	-0.0025	0.0229	0.0015