Quantifying Semantic Functional Specialization in the Brain Using Encoding Models of Natural Language

Jiaqi Chen[*]
MRC Laboratory of Molecular Biology
jc2130@cam.ac.uk

Richard J. Antonello[*]
Columbia University
rja2163@columbia.edu

Kaavya Chaparala
Johns Hopkins University
kchapar1@jh.edu

Coen Arrow
University of Western Australia
coen.arrow@research.uwa.edu.au

Nima Mesgarani
Columbia University
nima@ee.columbia.edu

Abstract

Although functional specialization in the brain - a phenomenon where different regions process different types of information - is well documented, we still lack precise mathematical methods with which to measure it. This work proposes a technique to quantify how brain regions respond to distinct categories of information. Using a topic encoding model, we identify brain regions that respond strongly to specific semantic categories while responding minimally to all others. We then use a language model to characterize the common themes across each region's preferred categories. Our technique successfully identifies previously known functionally selective regions and reveals consistent patterns across subjects while also highlighting new areas of high specialization worthy of further study.

1 Introduction

The theory of functional specialization states that different brain regions have evolved to process different types of information. This is apparent at a high level – for instance, the occipital lobe is heavily involved in processing visual information, whereas the temporal lobe is implicated in processing auditory stimuli. There is also evidence to support the theory at a lower level where smaller regions of interest (ROIs) are active in processing even more specific information. Examples of such ROIs include fusiform face area (FFA), which is selective for facial features, or places in parahippocampal place area (PPA). These discoveries have relied on 'contrast' studies that observe how brain regions respond to specific categories. While successful contrast studies have been influential, this approach also depends heavily on educated guesses about where in the brain to look and what to look for.

In this study, we propose an intuitive technique to identify and quantify functional specialization across the brain. We show that our method can correctly identify ROIs previously observed to have high semantic functional specialization, and that it can additionally identify several new ROIs with high functional selectivity throughout cortex for further study. We further demonstrate that this method can be used to explicitly recover known selectivity properties of well-documented regions such as the location-selective retrosplenial cortex (RSC) and the body-selective extrastriate body area (EBA), as well as for its newly-proposed regions. We additionally explore the degree to which our method generalizes across individuals, demonstrating that our observations are largely consistent across three subjects.

2 Related Work

Functional specialization has been studied in settings like vision (Kamps et al., 2016; Julian et al., 2016; Taylor et al., 2007; Calvo-Merino et al., 2010; Leibo et al., 2015; Saleem et al., 2018; Howard et al., 1996), language (Fedorenko et al., 2011), auditory processing (Perani et al., 2010; Tervaniemi et al., 1999), and motor function (Wilson et al., 2014). Previous work identifying areas of high functional specialization typically focuses on locating lateral asymmetries (Wang et al., 2014; Zilles and Amunts, 2015). However regions can be functionally specialized without being asymmetric, as is the case for regions like retrosplenial cortex (Mitchell et al., 2018; Burles et al., 2017), parahippocampal place area (Epstein and Kanwisher, 1998; Epstein, 2005), occipital place area (Kamps et al., 2016; Dilks et al., 2013), and extrastriate body area (Astafiev et al., 2004).

Researchers have also uncovered more direct links between semantics and brain activity by developing encoding models to predict neural responses

[*]Equal contribution.

Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, pages 77–90
May 3, 2025 ©2025 Association for Computational Linguistics

from continuous linguistic features. Mitchell et al. (2008) showed that different semantic categories (e.g., tools versus animals) elicit distinct spatial activation patterns in the brain, and a model trained on corpus-derived semantic features could forecast fMRI responses to previously unseen words. Later investigations built on this approach by applying distributed semantic representations to more complex, real-world language inputs (Huth et al., 2016; Jain and Huth, 2018; Caucheteux et al., 2023; Antonello et al., 2023). Utilizing high-dimensional word embeddings or semantic spaces derived from modern LLMs, these encoding models can capture brain responses to entire sentences and stories and generalize to numerous concepts. In this work, we build upon recent studies that use encoding models to generate and test interpretable hypotheses about semantic selectivity in the brain (Singh et al., 2023; Antonello et al., 2024).

3 Methods

3.1 fMRI Data

We used publicly available functional magnetic resonance imaging (fMRI) data collected from 3 human subjects as they listened to 20 hours of English language podcast stories over Sensimetrics S14 headphones. The stories came from podcasts such as *The Moth Radio Hour*, *Modern Love*, and *The Anthropocene Reviewed*. Each 10-15 minute story was played during a separate scan. Subjects were not asked to make any responses, but simply to listen attentively to the stories. For encoding model training, each subject listened to roughly 95 different stories, giving 20 hours of data across 20 scanning sessions, or a total of ~33,000 datapoints for each voxel in the brain.

MRI data were collected on a 3T Siemens Skyra scanner at The University of Texas at Austin Biomedical Imaging Center using a 64-channel Siemens volume coil. Functional scans were collected using a gradient echo EPI sequence with repetition time (TR) = 2.00 s, echo time (TE) = 30.8 ms, flip angle = 71°, multi-band factor (simultaneous multi-slice) = 2, voxel size = 2.6mm x 2.6mm x 2.6mm (slice thickness = 2.6mm), matrix size = 84x84, and field of view = 220 mm. Anatomical data were collected using a T1-weighted multi-echo MP-RAGE sequence with voxel size = 1mm x 1mm x 1mm.

In addition to motion correction and co-registration (LeBel et al., 2022), low-frequency voxel response drift was identified using a 2nd order Savitzky-Golay filter with a 120 second window and then subtracted from the signal. The mean response for each voxel was subtracted and the remaining response was scaled to have unit variance.

All subjects were healthy and had normal hearing. The experimental protocol was approved by the Institutional Review Board at The University of Texas, Austin. Written informed consent was obtained from all subjects.

3.2 Topic Encoding Model

A topic model was pre-trained on the entire story data. Given a list of word sequences for every two-second interval of the podcasts, we used sliding windows of [2, 4, . . . , 20] seconds. We trained 10 separate topic models on a different sliding window length each and then merged the topics to yield the final topic model, which had $T_n = 463$ topics. The multi-scale topic model was used to ensure the topic model generalized well across varied semantic timescales. In total, this training took $\sim$50 CPU node-hours.

Each model was based on the BERTopic technique (Grootendorst, 2022). Each string was embedded using sentence embedding model "all-MiniLM-L6-v2". Uniform Manifold Approximation and Projection (UMAP) was used to reduce the embedding dimension. The reduced embeddings were then clustered using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (McInnes et al., 2017).

Each cluster is associated with the the initial strings from each of the two-second intervals represented in the cluster. We manipulate these strings to extract a common semantic theme to describe the cluster. Initially, a bag-of-words is generated for each topic. The bags-of-words are then all concatenated into a single string and further reduced with class-based Term-Frequency Inverse-Document-Frequency (c-TF-IDF). The final representation reduction used KeyBERT(Grootendorst, 2020). The resultant meta-topics provide each cluster with a semantic theme common to all of its members.

The trained topic model encodes the story data, annotating the dataset using the general encoding methodology described in Antonello et al. (2021). Each 16 word phrase from the story dataset was fed through the topic model and was scored according to how likely each phrase is a member of each topic. The context length of 16 words was selected to bal-

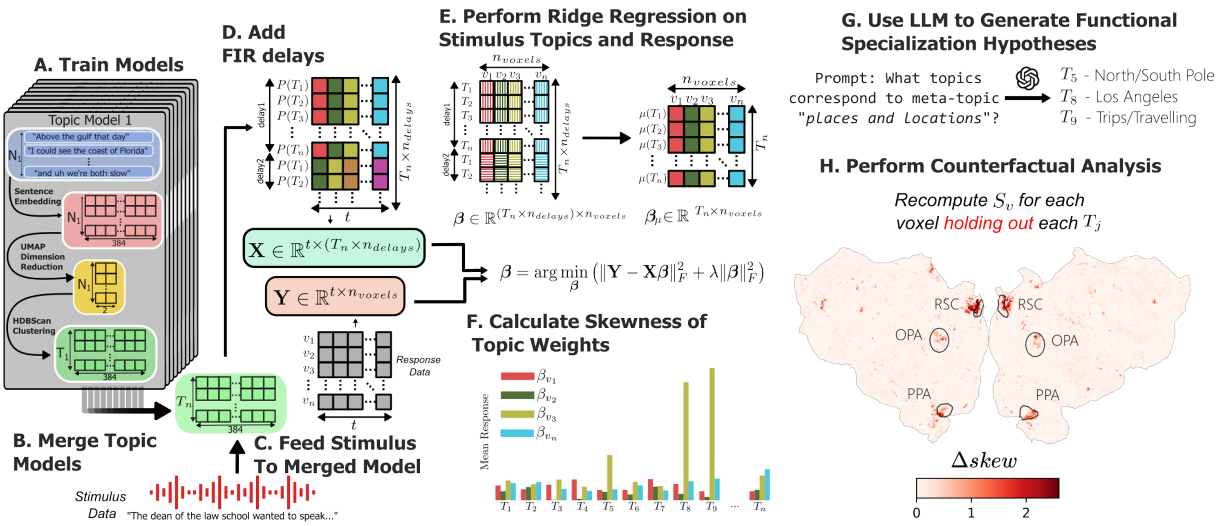

Figure 1: **Functional Selectivity Detection Pipeline**: Functional selectivity for a given meta-topic is generated using an automated pipeline. (A) 10 topic models are trained on strings of varying lengths derived from the stimulus dataset using a hierarchical clustering algorithm. (B) The topics generated are merged via cosine similarity, resulting in a single final topic model of T_n topics. (C) Story segments are fed into the newly generated merged topic model to build topic embeddings for the stimulus. (D) FIR delays are added to the topic embeddings, yielding our final stimulus matrix X. The BOLD response is temporally aligned with the time-delayed stimulus to produce our response matrix Y. (E) Bootstrapped ridge regression is used to generate weights β that map the stimulus to the voxelwise response. (F) We calculate the Pearson skewness of each voxel to find voxels with high functional selectivity. (G) An LLM is prompted to automatically determine which topics from the merged topic model correspond to a chosen meta-topic. (H) A counterfactual analysis is performed, where we observe which functionally selective voxels have their Pearson skewness reduced when the voxels from the corresponding meta-topic are excluded. The final flatmap shows the resulting drop in skewness for the "Places" meta-topic, demonstrating that the method correctly identifies RSC, OPA, and PPA as places-selective regions.

ance good performance of the model with a need to keep the topic labels relatively contemporaneous with the immediate content. These topic probabilities were then used as features for the encoding model. These features were then downsampled using Lanczos downsampling and finite impulse response (FIR) delays of 2, 4, 6 and 8 seconds were applied to model the hemodynamic response function (HRF) of the BOLD signal. A linear projection from these downsampled, time-delayed features to the measured BOLD response was then trained using bootstrapped ridge regression. That is, let X be the stimulus features derived from our topic model and let Y_v be the measured BOLD response for a given voxel v. For each voxel, we found linear weights β_v by optimizing

$$\min_{\beta_v}(Y_v - X\beta_v) + \lambda||\beta_v||_2 \qquad (1)$$

where $||\cdot||_2$ denotes the L_2-norm and λ is a regularization parameter.

3.3 Skewness as a Marker of Functional Selectivity

With β_v computed for every voxel, we used it to determine which brain voxels are highly selective for a small number of topics. To do this, we measured the voxelwise Pearson skewness of β_v. Intuitively, Pearson skewness is high when most of the weight in β_v is allocated to a small number of its elements. These highly-skew weights are more likely to be driven by the existence or nonexistence of a narrow number of topics in a given context. As functional selectivity is just the property of being selective for a narrow number of topics, high-skew voxels are by definition highly functionally-selective.

With highly-selective voxels identified, we segmented them into local clusters based on their cortical proximity. Contiguous sets of voxels of high skewness were grouped into proposed ROIs. For a given region $\mathcal{R}$, the weights β_v corresponding to the voxels in that region are then averaged and we observe topics with the highest weights in the averaged $\beta_{\mathcal{R}}$. If those topics shared a semantic category (for instance, if they are all "number words"),

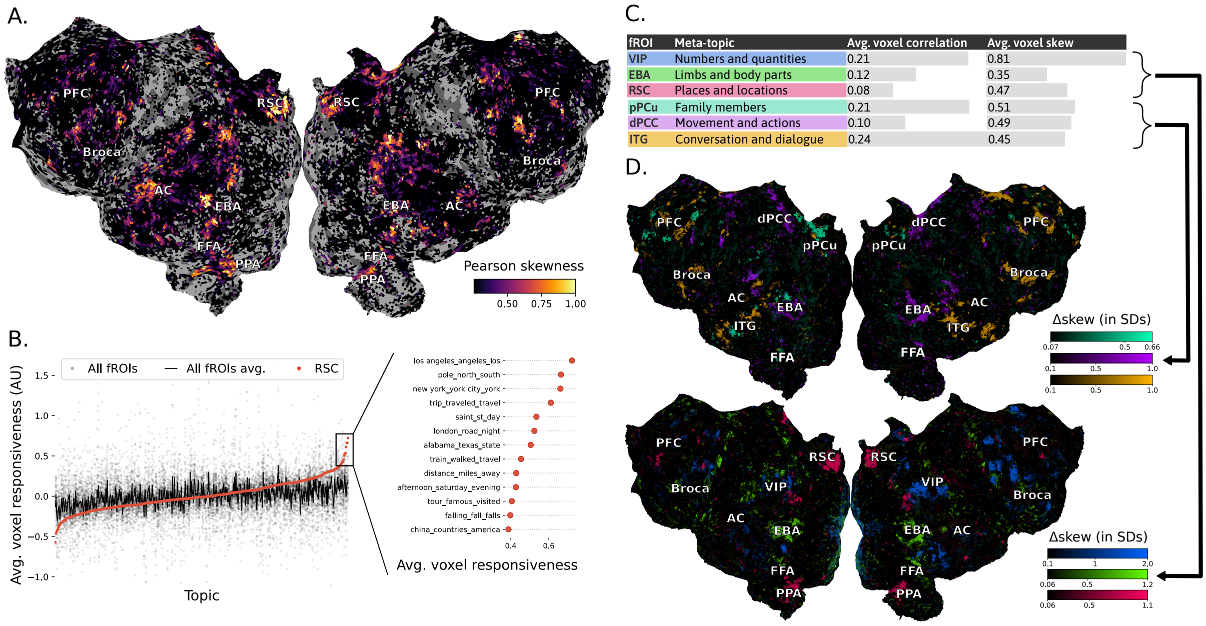

Figure 2: **Analyzing Functional Selectivity** (A) A cortical flatmap of voxels with high functional-selectivity using our skewness metric. The ventral visual stream has the highest overall functional-selectivity, whereas the frontal lobe has comparatively lower selectivity. (B) Individual topic responsiveness, according to our linear mapping β is visualized for the retrosplenial cortex. Most of the skewness in the distribution is derived from topics relating to location or time. Responses to these topics are substantially higher for RSC than other semantic categories. (C) Average voxel correlations and skewnesses for several fROIs are shown. Corresponding meta-topics are determined via post-hoc analysis of the most prominent topics. (D) Regions selective for the assorted meta-topics are visualized according to their relative selectivity. Voxels are colored according to how much omitting that set of meta-topics reduces overall skewness in that voxel. *Bottom:* Visualizes more novel, more poorly understood selectivities derived from our method. Meta-topics in (D) are colored according to the legend in (C).
Top: Visualizes well-understood semantic functional selectivities.

we concluded that the corresponding ROI is functionally selective for that category.

To map the putative functional organization of a semantic category or meta-topic, we prompted an LLM (GPT-4) (OpenAI, 2023) to select the top 6% of topics from the list of generated topics that semantically aligned most with the provided meta-topic. These output topics were checked for correctness and the top 20 (~4%) topics were selected. Voxel-wise, Z-scored Pearson skewness was then recalculated with and without the chosen topics. Since the average voxel had a z-scored skewness of zero, voxels selective for the meta-topic have high skew magnitudes when the meta-topic topics are included, but regress to zero once the topics were removed. The difference between the z-scored skews is the number of standard deviations from the mean skew that the meta-topic provides for the voxel.

We define the voxel-level Pearson skewness for any set of topic indices $\Omega \subseteq \{1, \dots, T_n\}$ as

$$s_v(\Omega) = \frac{\sum_{i\in\Omega}\left(\beta_{v,i} - \bar{\beta}_v(\Omega)\right)^3}{\left(\sum_{i\in\Omega}\left(\beta_{v,i} - \bar{\beta}_v(\Omega)\right)^2\right)^{3/2}} \qquad (2)$$

where $\bar{\beta}_v(\Omega) = \frac{1}{|\Omega|}\sum_{i\in\Omega}\beta_{v,i}$.

After z-scoring these skewness values across voxels (so that the mean voxel has zero z-scored skewness), the contribution of a chosen meta-topic $\mathcal{M}$ is then

$$\Delta\text{skew}_v(\mathcal{M}) = Z\big(s_v(\Omega_{\text{all}})\big) - Z\big(s_v(\Omega_{\text{all}}\setminus\mathcal{M})\big), \qquad (3)$$

where $\Omega_{\text{all}} = \{1, \dots, T_n\}$ is the full set of topics.

This skewness differential is used to measure the degree to which voxel v is selective for the selected meta-topic. If a voxel has high-initial skewness, and then that skewness is substantially reduced when topics from a given meta-topic are excluded, we conclude that that voxel is functionally selective for the associated meta-topic. The full pipeline for functional selectivity detection is depicted in **Figure 1**.

3.4 Analyzing Meta-topic Specificity

Following the application of our functional selectivity pipeline to our data, we observed a small number of regions with apparently strong semantic

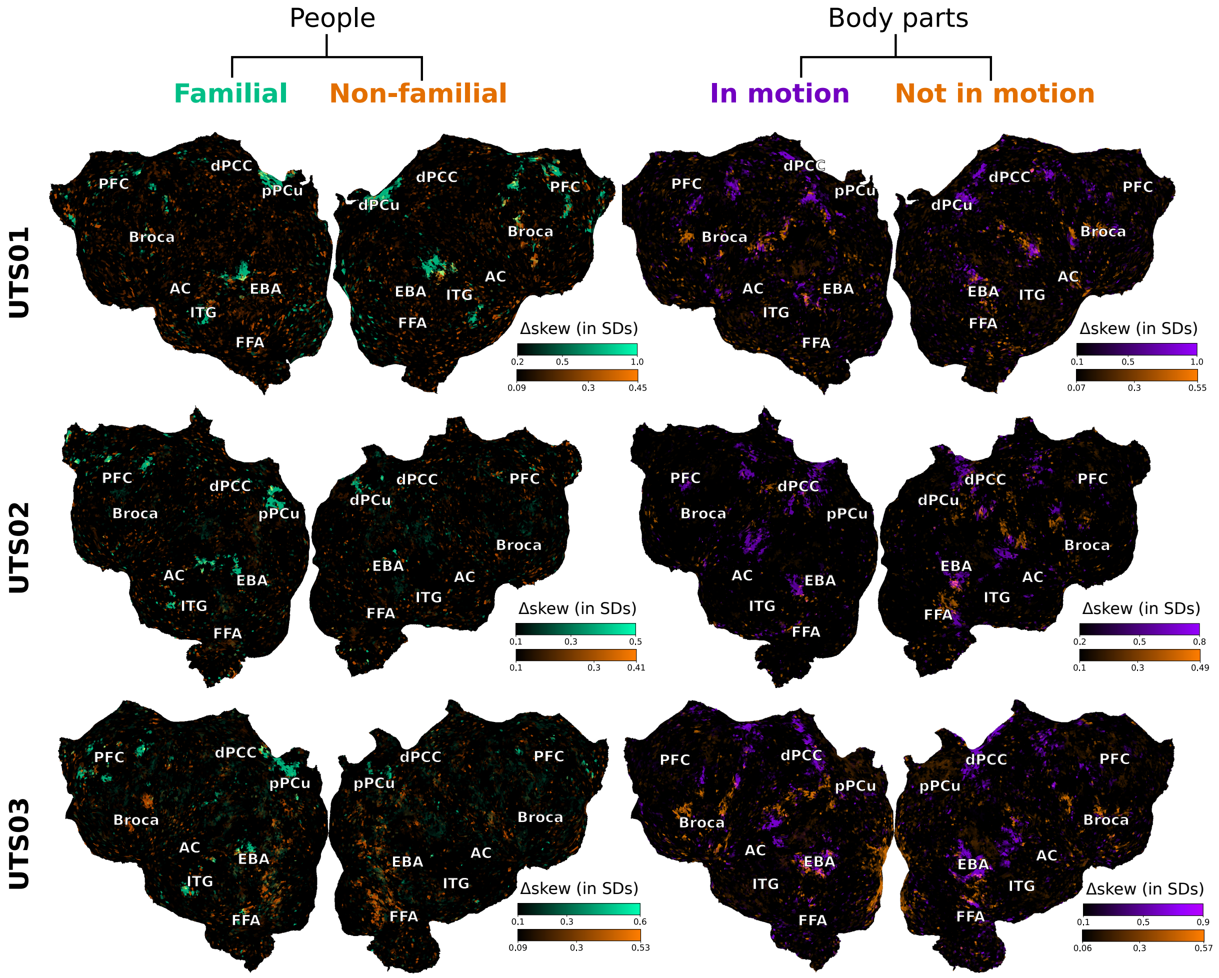

Figure 3: **Investigating Meta-topic Precision for Novel Regions** For each of our three subjects (*Top:* UTS01, *Center:* UTS02, *Bottom:* UTS03), we visualize the meta-topic selectivity of familial people vs. non-familial people (*Left*) and body parts in motion vs. not in motion (*Right*), In each case, we find that the selectivity of the corresponding region is precisely captured by the chosen meta-topic (i.e. familial people for pPCu; moving body parts for dPCC) and not its antithesis.

functional selectivity, according to our metric, that did not appear to have a strong basis in prior literature. To determine the meta-topic specificity these novel ROIs, we provided GPT-4 with a theme one level broader than the ROI's proposed meta-topic and analyzed the previously missed topics. For example, if an ROI is found selective for family members, we prompted GPT-4 to select the top 6% of topics that are most aligned with people but do not reference family. Analysis on the skewness differential was then repeated for the top 20 topics of this new, broader meta-topic. This process ensures that the meta-topic selected possesses the highest level of granularity for which that ROI is functionally selective.

3.5 Measuring Anatomical Consistency

To validate the generality of the anatomical observations made from the topic selectivity encoding models, we follow a methodology similar to that of Huth et al. (2016) in observing the semantic tiling of the cortex with respect to our topic space. In particular, we perform PCA along the topic axis of our linear mapping for one subject (UTS03) to get a set of orthogonal principal components that maximally explain the variance along that axis. We then project these components to the voxel space for every subject, by computing the dot product of the topic components with the specific linear encoding weights for that subject. The anatomical alignment of the resulting projection between subjects determines the degree to which observations derived from our topic encoding models are population trends.

4 Results

Figure 2 shows the results of applying our functional selectivity mapping protocol to a single subject from our fMRI dataset. Additional results for the other two subjects in our study are presented in Appendix A. Results are highly consistent across the subjects in our dataset, suggesting that the functional selectivities described here are population-level trends.

Figure 2a shows the voxels described as functionally selective according to our topic skewness metric. We find that many regions previously identified as functionally selective are correctly labeled as such by our skewness metric. The highest functional selectivity is observed along a band of the higher ventral visual stream, which includes regions like occipital place area, extrastriate body area, and borders interparietal sulcus. We observed comparatively smaller amounts of functional selectivity in regions outside the ventral visual stream, such as the prefrontal cortex. This may suggest that the presence of local visual representations is a strong driver of selectivity, supporting grounding theories of cognition (Barsalou, 2008) that suggest that neural representations are "grounded" in sensorimotor information. Additionally, the result replicates prior work suggesting a visio-semantic alignment that occurs at the border of visual cortex (Popham et al., 2021).

Figure 2b demonstrates the process of isolating the functional selectivity of one such ROI through the example of a set of voxels located in RSC. We replicate the established result that this ROI is highly selective for "place semantics", as suggested by prior literature (Mitchell et al., 2018). Most topics with the highest weight on these voxels are semantically associated with locations, travel, or temporal concepts. The top three topics for this set of voxels (*New York, North/South Pole, Los Angeles*) are all prominent geographical locations. We additionally note a further time-semantics component to some of the most selective topics (e.g. *afternoon, night, day*), suggesting that RSC is further implicated in the processing of temporal information, not just spatial information. Additional analysis of the temporal profiles of the topics suggests that this is not due to place- and time-related topics co-occurring naturally in the stimuli, but instead due to separate and independent effects for both meta-topics. (Appendix Figure 7 and Appendix Table 1).

Figure 2c further shows this process applied to six selected fROIs, with the meta-topics associated with their functional selectivities. Average predictive performance and voxel skew for the voxels in these regions are also shown. All selected regions have high skewness and most have high prediction performance relative to the average cortical voxel. Meta-topic descriptions are consistent with prior literature, and we correctly identify the functional selectivity of regions such as extrastriate body area (EBA), ventral interparietal area (VIP) and parahippocampal place area (PPA). It is noteworthy that while selectivity in areas like EBA has been established previously using visual contrasts, our method replicates these findings purely from linguistic input, further supporting those claims in a modality distinct from prior work.

Figure 2d shows the cortex-wide meta-topic selectivity for the concepts of three previously-established functionally-selective regions: RSC, EBA and VIP. We find several other less well-described functional selectivity regions, such as an area near inferior temporal gyrus (ITG) cortex that is selective for "conversation" words, an area in posterior precuneus (pPCu) that is selective for "family"-related words, and an area that is selective for words describing movement or physical actions of body parts, which is located in dorsal posterior cingulate cortex (dPCC). Voxels are colored according to how much that voxel's skewness would change if the corresponding meta-topic's weights were removed.

Given that we found the existence of these newer regions surprising owing to their unusual specificity, we examined the degree to which the descriptors of newly described regions were precise. For the posterior precuneus and dorsal cingulate ROIs, **Figure 3** looks at the effect of subtly altering the meta-topic we have associated with each region. Remarkably, we find that for pPCu, topic selectivity is no longer observed in the region when we choose a meta-topic of "non-familial" people. Similarly, the dorsal cingulate is not selective for body part actions that do not involve active movement. These effects are consistent across all 3 subjects in our study, further supporting the claim that these descriptors are indeed accurate summaries of the semantic functional selectivity of these regions. We believe these results strongly warrant further investigation given their surprising consistency and specificity.

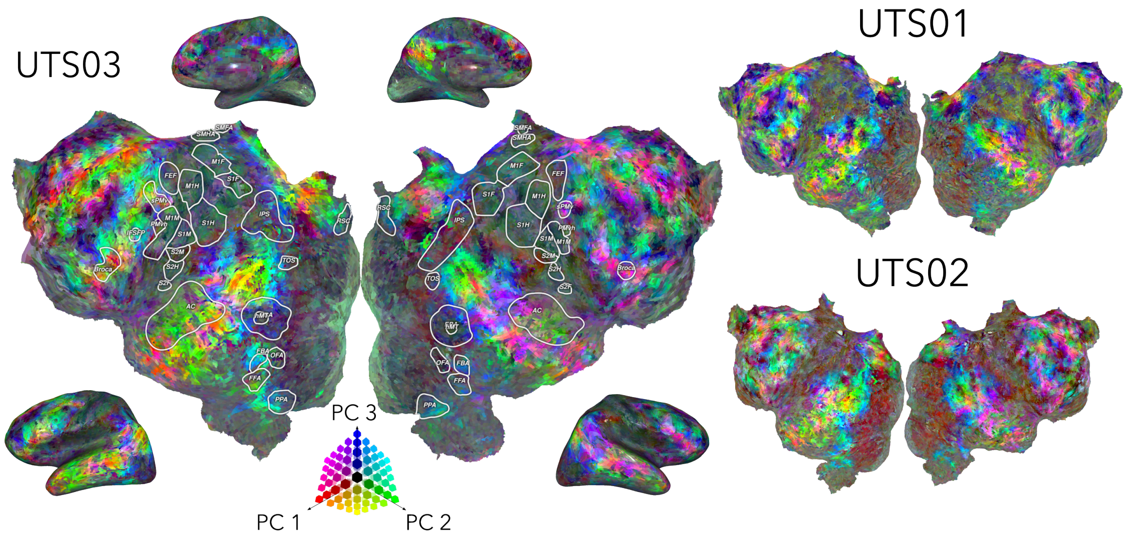

Figure 4: **Population-level Topic Selectivity** For each of our three subjects, the top 3 principal components of the topic model space for UTS03 are computed and then projected into the respective voxel space using the linear encoding model weights. The final projection is plotted, where each color channel denotes a single principal component from the topic space (*Red:* PC1, *Green:* PC2, *Blue:* PC3). We see largely consistent anatomical alignment across subjects, suggesting that topic-level selectivity generalizes to population-level trends.

Finally, we analyzed whether the weights from our topic encoding models were largely subject-specific, or generalized across the population. **Figure 4** shows the result of the principal component analysis of our topic encoding models. We see strong anatomical alignment across subjects, with especially prominent laterality patterns. For example, linear combination of PC1 and PC2 (*yellow*) is more observable in left temporal cortex, whereas a linear combination of PC1 and PC3 (*pink*) is more observable in right temporal cortex. Most importantly, this strong anatomical alignment across three subjects provides good support for the claim that the topic encoding models are largely consistent across individuals and are therefore not heavily influenced by subject-level differences.

5 Discussion and Limitations

Unlike today's computers, which are no more than collections of billions of identical and functionally equivalent transistors, the brain is no computational monolith. Despite this well-known fact, remarkably little effort has gone into designing methods to automatically detect and characterize this functional selectivity, especially in the realm of language semantics. To this day, functional selectivity is primarily analyzed through painstaking and tedious "contrast studies" in which subjects are exposed to carefully curated experiments, in order to narrow down the functional selectivity of a region. Here, we show that by utilizing modern machine learning techniques, we can detect and analyze functional selectivity in a vacuum.

Replicating prior studies (Mitchell et al., 2018; Burles et al., 2017; Astafiev et al., 2004), our results show many regions of the brain are highly selective for specific semantic categories, such as places, conversations, or body parts. We further explore evidence for functional selectivity of less well-understood regions like posterior precuneus and dorsal posterior cingulate cortex, showing that they are selective for the highly specific concepts of family members and movement-based actions respectively. We find this surprising, but are able to show that this selectivity is consistent across subjects and actually requires this level of specificity. We show that functional selectivity is most heavily distributed along the ventral visual stream, but is also present to a lesser degree in areas such as prefrontal cortex. The functional selectivity we detect tends to be more biased toward concrete concepts over abstract ones, suggesting that more abstract concepts are less likely to have uniquely specialized regions. Nevertheless, this ongoing work has several limitations. Firstly, the proximity-based process of clustering voxels into fROIs is still rudimentary and could be supplemented with a more nuanced approach that directly takes into account similarity in voxel weights. Next, the meta-topic classification that is currently performed by an LLM could be subject to additional validation on its agreement with human classification. Further work should also be pursued into understanding individual subject-level differences in functional selectivity to determine the degree to which these observations hold across the population. The relative data-efficiency of our method could provide a more fruitful perspective into these differences across subjects.

What can functional selectivity tell us about the nature of human intelligence? We often find that, outside of the ventral visual stream, most functional selectivity is closely related to the non-semantic role of adjacent regions. For example, the semantic selectivities of dPCu (movement and actions) and ITG (conversations) are closely associated with their non-semantic roles (motor planning and auditory processing). We find such "functional coincidences" to be persuasive evidence in support of cognitive grounding, the notion that cognitive representations are "ultimately grounded in bodily, affective, perceptual, and motor processes" (Pezzulo et al., 2013), rather than "computation on amodal symbols in a modular system" (Barsalou, 2008). In a model of intelligence based on grounded cognition, functionally-selective regions would likely benefit from their proximity to areas specialized in related low-level processes. Further research into the mapping of functional selectivity could one day help to reveal the underlying organizational drivers of cortical structure.

Acknowledgments

This work is based on work done during the 2024 Telluride Neuromorphic Cognition Engineering Workshop, sponsored by the NSF. Additional funding was provided by the National Institutes of Health (NIH-NIDCD), the Dingwall Foundation, and a grant from Marie Josee and Henry R. Kravis. We thank Alexander Huth and Nicole Dumont for their insightful feedback and assistance during this project.

References

Richard Antonello, Chandan Singh, Shailee Jain, Aliyah Hsu, Jianfeng Gao, Bin Yu, and Alexander Huth. 2024. A generative framework to bridge data-driven models and scientific theories in language neuroscience. *arXiv preprint arXiv:2410.00812*.

Richard Antonello, Javier S Turek, Vy Vo, and Alexander Huth. 2021. Low-dimensional structure in the space of language representations is reflected in brain responses. *Advances in Neural Information Processing Systems*, 34.

Richard Antonello, Aditya Vaidya, and Alexander G. Huth. 2023. Scaling laws for language encoding models in fmri. *Preprint*, arXiv:2305.11863.

Serguei V Astafiev, Christine M Stanley, Gordon L Shulman, and Maurizio Corbetta. 2004. Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature neuroscience*, 7(5):542–548.

Lawrence W Barsalou. 2008. Grounded cognition. *Annu. Rev. Psychol.*, 59(1):617–645.

Ford Burles, Edward Slone, and Giuseppe Iaria. 2017. Dorso-medial and ventro-lateral functional specialization of the human retrosplenial complex in spatial updating and orienting. *Brain structure & function*, 222:1481–1493.

Beatriz Calvo-Merino, Cosimo Urgesi, Guido Orgs, Salvatore M Aglioti, and Patrick Haggard. 2010. Extrastriate body area underlies aesthetic evaluation of body stimuli. *Experimental brain research*, 204:447–456.

Charlotte Caucheteux, Alexandre Gramfort, and Jean-Rémi King. 2023. Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nature Human Behaviour*, pages 1–12.

Daniel D. Dilks, Joshua B. Julian, Alexander M. Paunov, and Nancy Kanwisher. 2013. The occipital place area is causally and selectively involved in scene perception. *Journal of Neuroscience*, 33:1331–1336.

Russel A. Epstein. 2005. The cortical basis of visual scene processing. *Visual Cognition*, 12:954–978.

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

Evelina Fedorenko, Michael K Behr, and Nancy Kanwisher. 2011. Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, 108(39):16428–16433.

Maarten Grootendorst. 2020. Keybert: Minimal keyword extraction with bert.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

R.J. Howard, M. Brammer, I. Wright, P.W. Woodruff, E.T. Bullmore, and S. Zeki. 1996. A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. *Current Biology*, page 1015–1019.

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

Shailee Jain and Alexander Huth. 2018. Incorporating context into language encoding models for fmri. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.

Joshua B Julian, Jack Ryan, Roy H Hamilton, and Russell A Epstein. 2016. The occipital place area is causally involved in representing environmental boundaries during navigation. *Current Biology*, 26(8):1104–1109.

Frederik S Kamps, Joshua B Julian, Jonas Kubilius, Nancy Kanwisher, and Daniel D Dilks. 2016. The occipital place area represents the local elements of scenes. *Neuroimage*, 132:417–424.

Amanda LeBel, Lauren Wagner, Shailee Jain, Aneesh Adhikari-Desai, Bhavin Gupta, Allyson Morgenthal, Jerry Tang, Lixiang Xu, and Alexander G Huth. 2022. A natural language fmri dataset for voxelwise encoding models. *bioRxiv*, pages 2022–09.

Joel Z Leibo, Qianli Liao, Fabio Anselmi, and Tomaso Poggio. 2015. The invariance hypothesis implies domain-specific regions in visual cortex. *PLoS Computational Biology*, 10.

Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205.

Anna S Mitchell, Rafal Czajkowski, Ningyu Zhang, and Andrew J D Jeffery, Kate adn Nelson. 2018. Retrosplenial cortex and its role in spatial cognition. *Brain and neuroscience advances*, 2.

Tom M Mitchell, Svetlana V Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L Malave, Robert A Mason, and Marcel Adam Just. 2008. Predicting human brain activity associated with the meanings of nouns. *science*, 320(5880):1191–1195.

OpenAI. 2023. Gpt-4 system card.

Daniela Perani, Maria Cristina Saccuman, Paola Scifo, Danilo Spada, Guido Andreolli, Rosanna Rovelli, Cristina Baldoli, and Stefan Koelsch. 2010. Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences*, 107(10):4758–4763.

Giovanni Pezzulo, Lawrence W Barsalou, Angelo Cangelosi, Martin H Fischer, Ken McRae, and Michael J Spivey. 2013. Computational grounded cognition: a new alliance between grounded cognition and computational modeling. *Frontiers in psychology*, 3:612.

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

Aman B. Saleem, E. Mika Diamanti, Julien Fournier, Kenneth D. Harris, and Matteo Carandini. 2018. Coherent encoding of subjective spatial position in visual cortex and hippocampus. *Nature*, 562:124–127.

Chandan Singh, Aliyah R Hsu, Richard Antonello, Shailee Jain, Alexander G Huth, Bin Yu, and Jianfeng Gao. 2023. Explaining black box text modules in natural language with language models. *arXiv preprint arXiv:2305.09863*.

John C Taylor, Alison J Wiggett, and Paul E Downing. 2007. Functional mri analysis of body and body part representations in the extrastriate and fusiform body areas. *Journal of neurophysiology*, 98(3):1626–1633.

Mari Tervaniemi, Anu Kujala, Kimmo Alho, Juha Virtanen, Risto J Ilmoniemi, and Risto Näätänen. 1999. Functional specialization of the human auditory cortex in processing phonetic and musical sounds: a magnetoencephalographic (meg) study. *Neuroimage*, 9(3):330–336.

Danhong Wang, Randy L Buckner, and Hesheng Liu. 2014. Functional specialization in the human brain estimated by intrinsic hemispheric interaction. *Journal of Neuroscience*, 34(37):12341–12352.

Tony W Wilson, Max J Kurz, and David J Arpin. 2014. Functional specialization within the supplementary motor area: a fnirs study of bimanual coordination. *NeuroImage*, 85:445–450.

Karl Zilles and Katrin Amunts. 2015. Anatomical basis for functional specialization. *fMRI: From nuclear spins to brain functions*, pages 27–66.

A Supplementary subject data

Here we reproduce Figure 2 for the other two subjects that we examined. We observe similar selectivity patterns throughout the cortex.

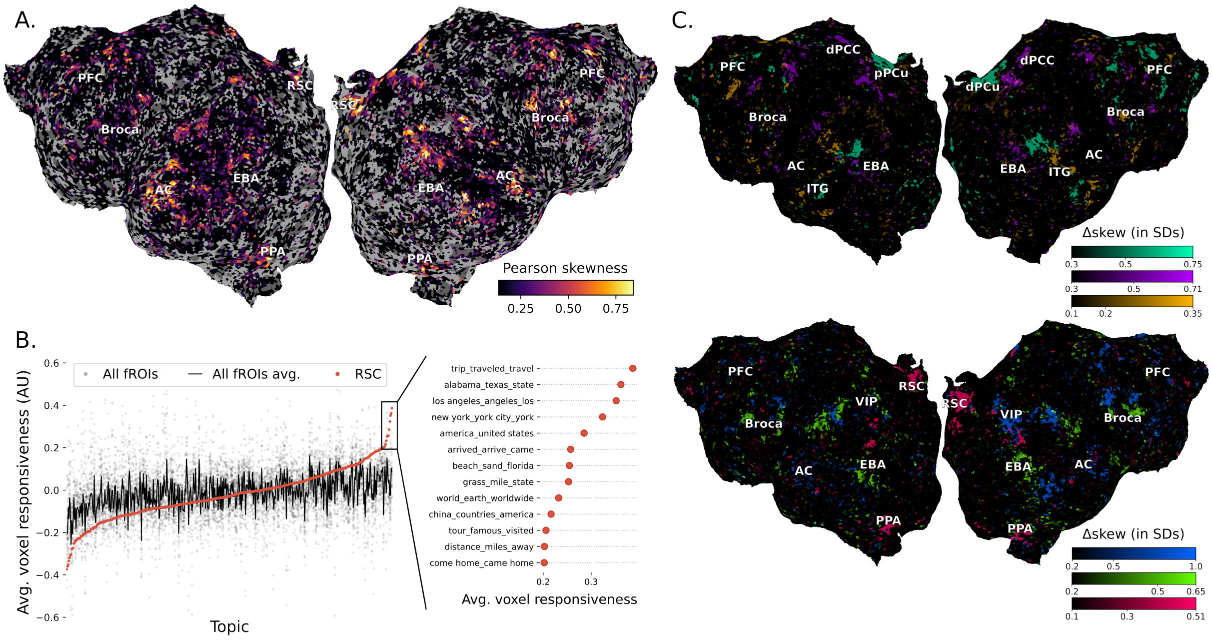

Figure 5: Replication of Figure 2 for UTS01. The same meta-topics and color key are used for (C) as in Figure 2D.

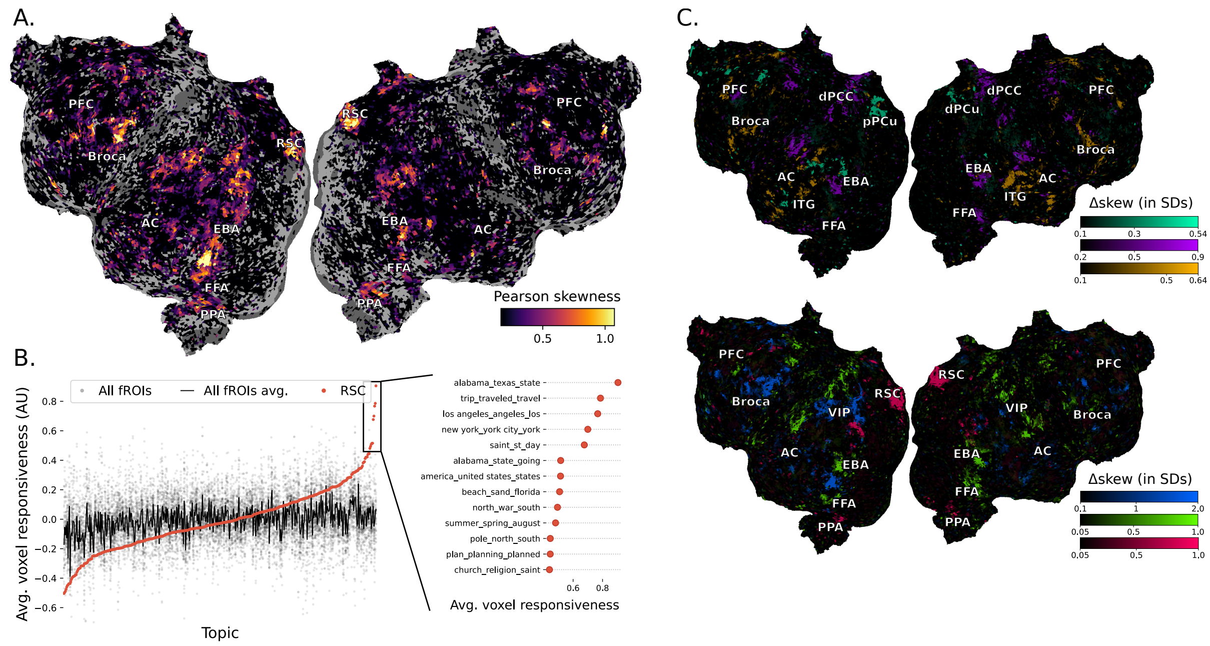

Figure 6: Replication of Figure 2 for UTS02. The same meta-topics and color key are used for (C) as in Figure 2D.

B Meta-topic data

The 20 topics that comprise each meta-topic are listed in Figure 8. This includes the six meta-topics that relate to an ROI (‘numbers and quantities’, ‘limbs and body parts’, ‘places and locations’, ‘family members’, ‘movement and actions’ and ‘conversation and dialogue’) and the two broader-level meta-topics that were used to determine ROI specificity in Section 3.4 (‘people excluding family’ and ‘body parts without movement’).

To test whether topics within a meta-topic tended to be predicted at similar timepoints in each story, we calculated pair-wise correlations between all 463 topics based on their probability scores per TR. To minimize the effect of noise from low-probability scores, only the top 10% of scores within each TR were used in this calculation while the bottom 90% were set to zero. Then, maximum-distance hierarchical clustering on pairwise correlations was used to group the topics into 102 clusters. Figure 7 shows these pair-wise correlations, with the topics reordered to visualize the clusters in their hierarchical order.

This data was used to determine whether the co-existence of more than one topic theme in an ROI’s apparent topic selectivity is likely due to an actual functional selectivity, or an artifact of topic co-expression in the same sentences in the story data. For example, the data suggests RSC to be indeed functionally selective for both temporal topics (cluster 22) and geographical topics (clusters 71 and 92), as the inter-cluster correlations for cluster 22 are significantly lower than the intra-cluster correlations (Figure 7 and Table 1).

Figure 8 shows the correlation matrix subsetted to display values relevant to the listed meta-topic topics only. The involvement of many different clusters within each meta-topic similarly suggests our results on ROI functional selectivity (Figures 2 and 3) to not be heavily dependent on topic co-expression.

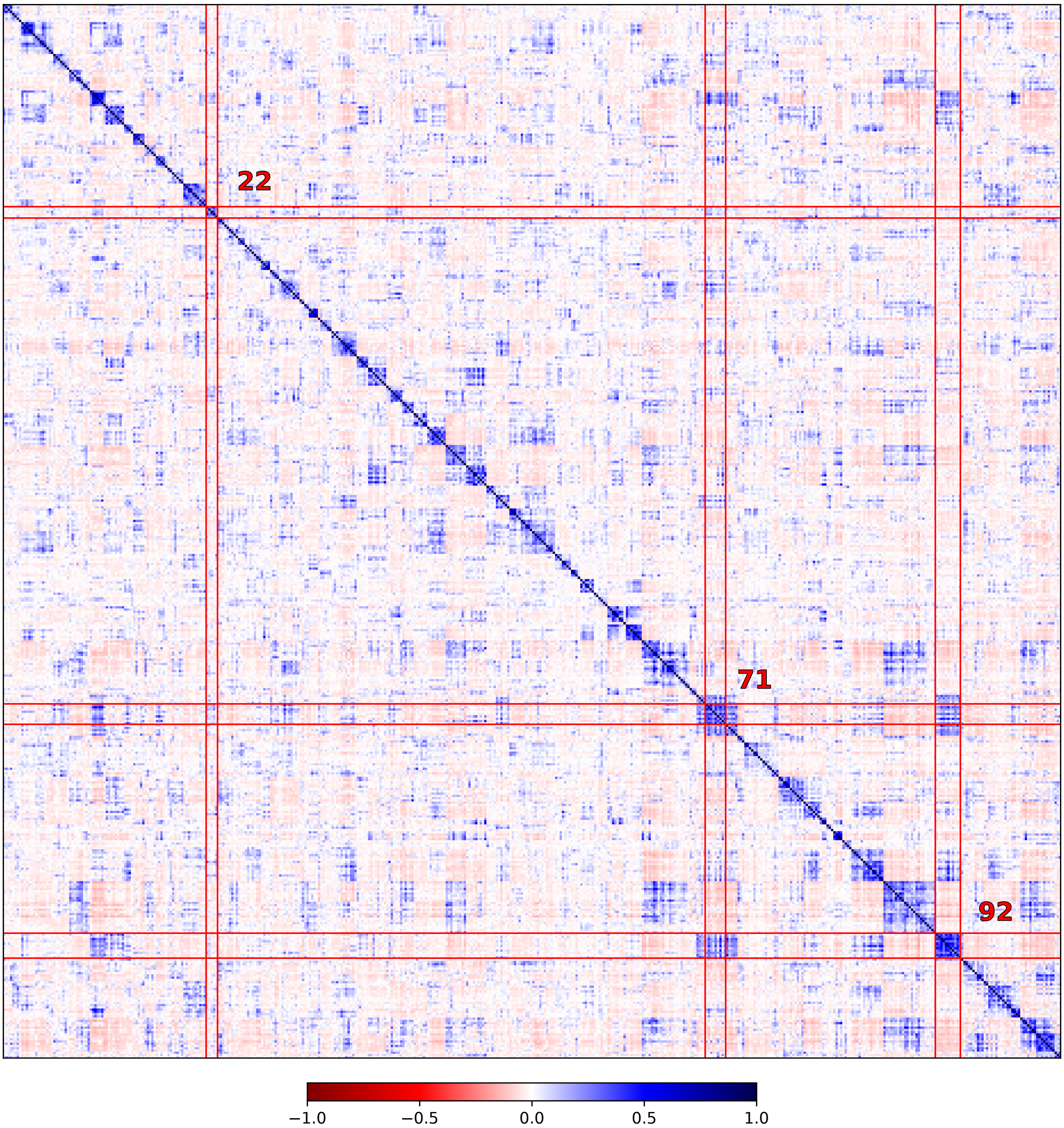

Figure 7: Biclustering results on the 463 topics of the final topic model. The x- and y-axis are symmetrical, with each row and column organized in hierarchical order following maximum-distance clustering. Clusters highlight topics that tend to occur near each other in time (i.e. occur within the same sentence) in the story data. Three clusters that contain the top RSC topics shown in Figure 2B have been highlighted in *red*, and the topics contained listed under Table 1.

Cluster	Topics
22	afternoon_saturday_evening
	lights_light_bright
	sleep_sleeping_slept
	wake_terrifying_right
	clock_morning_time
71	trip_traveled_travel
	come home_came home_home
	camp_park_meet
	london_road_night
	escort_father_family
	staying_stayed_stay
	ticket_tickets_flight
	funeral_celebrating_family
	train_walked_travel
92	new york_york city_york
	los angeles_angeles_los
	distance_miles_away
	alabama_texas_state
	north_war_south
	beach_sand_florida
	america_united states_states
	streets_east_street
	pole_north_south
	grass_mile_state
	alabama_state_going

Table 1: List of topics contained in the clusters labeled in Figure 7.

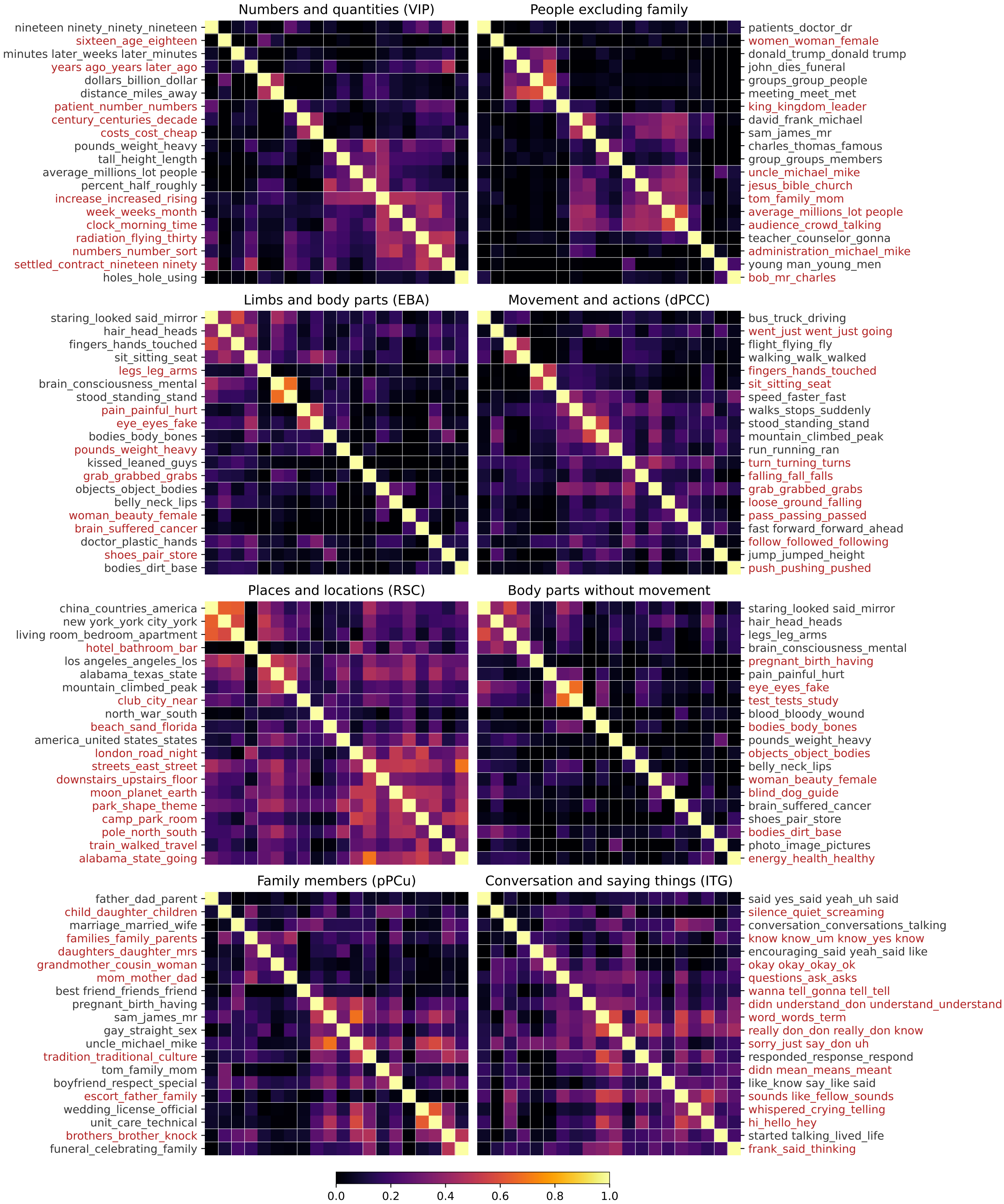

Figure 8: The topics comprising each meta-topic and their pairwise correlation scores. X-axis and y-axis labels are symmetrical. White borders and topic colors show the different cluster groups. Correlation scores were calculated by comparing the prediction profiles of each topic over all TRs.