Mapping Multimodal Conceptual Representations within the Lexical-Semantic Brain System

Emily Meschke (emily_meschke@berkeley.edu)
University of California, Berkeley
Berkeley, CA USA

Jack Gallant (gallant@berkeley.edu)
University of California, Berkeley
Berkeley, CA USA

The representation of modal (e.g., visual, auditory, motor) and multimodal conceptual information underlies people's ability to reason and communicate about the world. However, little is known about the form and spatial organization of multimodal conceptual representations. Some studies suggest that lexical-semantic representations, located within the broadly distributed lexical-semantic brain system, play an important role in linking conceptual representations across modalities. We therefore consider whether regions within the lexical-semantic system represent the same semantic information during narrative speech comprehension and movie watching. Eleven participants listened to narrative stories and watched compilations of silent short movie clips while their brain responses were recorded with fMRI. Two separate voxelwise encoding models were created for each participant: one to map lexical-semantic representations during narrative comprehension, and one to map visual-semantic representations during movie watching. Multimodal regions within the lexical-semantic system were then identified as regions where both models predict responses across the two experiments. This cross-experiment prediction analysis revealed a set of multimodal convergence zones that border the lexical- and visual-semantic systems, as well as a set of multimodal patches distributed within the larger lexical-semantic system. These results support the hypothesis that semantic representations within the lexical-semantic system are involved in linking conceptual representations across modalities, and in forming our conceptual understanding of the world.

Keywords: voxelwise encoding models; multimodal representations; language; semantic cognition

Introduction

The representation of conceptual knowledge has been an important topic of debate in cognitive neuroscience for several decades. Most theories agree that conceptual information is represented across both modal and multimodal systems (Damasio, 1989; Binder et al., 2009; Ralph et al., 2017). Modal systems include, but are not limited to, sensory systems that extract conceptual information from low-level sensory features, and motor systems that implement actions according to conceptual goals. However, there is less consensus about the form and spatial organization of multimodal systems.

Evidence suggests that the lexical-semantic system plays an important role in representing multimodal conceptual information (Fernandino et al., 2015; Tong et al., 2022). First, the lexical-semantic system covers a wide range of the cerebral cortex (Binder et al., 2009; Huth et al., 2016). Thus, this system has an efficient spatial organization for supporting convergence zones, or regions that interface locally with broadly distributed modal systems (Damasio, 1989; Devereux et al., 2013; Fairhall & Caramazza, 2013; Popham et al., 2021). Second, compared to sensory representations, lexical-semantic representations are compact. Thus, these representations are in an efficient form for linking rich sensory information to information in resource limited memory systems (Tulving, 1972). Therefore, we hypothesize that the broadly distributed lexical-semantic system plays a role in representing multimodal conceptual information.

Methods

To test the hypothesis that the lexical-semantic system represents multimodal conceptual information, we used voxelwise encoding models (Naselaris et al., 2011) to evaluate whether regions within the lexical-semantic system represent the same semantic information during two naturalistic experiments. In a lexical comprehension experiment, brain responses were recorded as 11 participants listened to naturally-spoken narrative stories (Huth et al., 2016). In a visual perception experiment, brain responses were recorded as the same 11 participants watched compilations of short movie clips (Nishimoto et al., 2011). For each experiment, low-level sensory and high-level semantic features were first extracted from the stimuli, and were then used to predict brain activity in each voxel of each participant. Then, for each participant, the lexical-semantic models trained on the narrative story data and the visual-semantic models trained on the short movie clip data were projected to the fsaverage6 template surface. Lastly, two group models were created by averaging the lexical- and visual-semantic models across participants, and the group models were used to predict each participants' held-out brain responses to the movies and stories. Regions that are well-predicted by both models contain multimodal conceptual representations (i.e., semantic representations that are relevant for both visual perception and lexical comprehension).

Results

Multimodal convergence zones between the visual- and lexical-semantic systems. We first identified regions where both the visual- and lexical-semantic models predict brain responses to the movie clips (Figure 1). Visual-semantic models predict responses across the mid- to high-order visual cortices, the intraparietal sulcus, and the frontal eye fields (FEF). Lexical-semantic models predict responses within the temporoparietal junction (TPJ), lateral temporal cortex (LTC), and medial parietal cortex (MPC). Both models predict responses in regions bordering the visual- and lexical-semantic systems, such as the parietal place area (PPA), extrastriate body area (EBA), and retrosplenial cortex (RSC). The spatial organization of these multimodal regions suggests that they reflect convergence zones between the semantic representations derived from vision and the semantic representations derived from language. Thus, the lexical-semantic system contains multimodal conceptual representations that reflect convergence between the visual- and lexical-semantic systems.

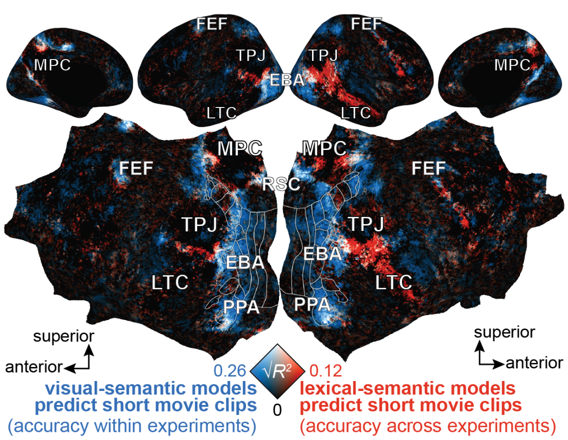

Figure 1: Brain responses to the movie clips were predicted using the visual-semantic models trained on the movie clips (well predicted regions in blue), and the lexical-semantic models trained on the stories (well predicted regions in red). Multimodal convergence zones that are well predicted by both the visual- and lexical-semantic models (in white) lie along the border between the visual- and lexical-semantic systems.

Multimodal patches within the lexical-semantic system. We then identified regions where both the lexical- and visual-semantic models predict brain responses to the narrative stories (Figure 2). Lexical-semantic models predict responses across a wide range of cortical regions, including TPJ, LTC, MPC, ventral prefrontal cortex (vPFC), and dorsal prefrontal cortex (dPFC). Visual-semantic models predict responses in a smaller subset of patches across these cortical regions. Unlike the previous analysis, the multimodal regions that are well predicted by both models are not spatially organized across a set of convergence zones between the visual- and lexical-semantic systems. Rather, these multimodal regions are broadly distributed across the larger lexical-semantic system. This spatial organization suggests that these multimodal regions reflect a subset of regions within the lexical-semantic system that process the semantic information recovered both from visual scenes and from language. Thus, the lexical-semantic system contains multimodal conceptual representations that are relevant for both visual scene perception and for language comprehension.

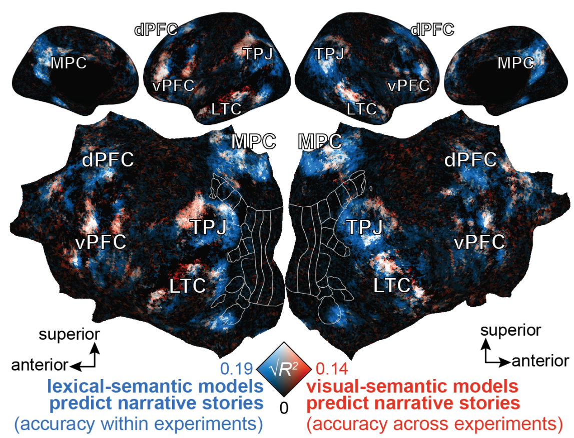

Figure 2: Brain responses to the narrative stories were predicted using the lexical-semantic models (in blue), and the visual-semantic models (in red). Multimodal regions that are well predicted by both lexical- and visual-semantic models (in white) are broadly distributed across the lexical-semantic system.

Conclusions

We showed that the lexical-semantic system contains multimodal representations. First, we identified multimodal convergence zones along the anterior border of the visual cortex, where visual-semantic representations converge with lexical-semantic representations. Second, we identified a subset of multimodal regions within the broader lexical-semantic system that represent semantic information derived from both vision and language. These results suggest that the same functional architecture that is used to represent conceptual information derived from language is used to represent conceptual information derived from sensory experiences. Thus, this work lends strong support to the hypothesis that semantic representations within the lexical-semantic system link conceptual representations across modalities, thereby underlying our conceptual understanding of the world.

Acknowledgments

This work was supported by the National Eye Institute (R01-EY031455) and the Office of Naval Research (N00014-20-1-2002).

References

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Comput.,* 1, 123–132.

Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *J. Neurosci.,* 33, 18906–18916.

Fairhall, S. L. & Caramazza, A. (2013). Brain regions that represent amodal conceptual knowledge. *J. Neurosci.,* 33, 10552–10558.

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., Conant, L. L., & Seidenberg, M. S. (2016). Concept representation reflects multimodal abstraction: a framework for embodied semantics. *Cerebral Cortex*, 26(5), 2018–2034.

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

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, *56*(2), 400–410.

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, 21(19), 1641–1646.

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.*, 18, 42–55.

Tong, J., Binder, J. R., Humphries, C., Mazurchuk, S., Conant, L. L., & Fernandino, L. (2022). A distributed network for multimodal experiential representation of concepts. *J. Neurosci.*, 42(37), 7121–7130.