

# Time-resolved brain activation patterns reveal hierarchical representations of scene grammar when viewing isolated objects

Poster Presentation 56.451: Tuesday, May 21, 2024, 2:45 – 6:45 pm, Pavilion  
Session: [Object Recognition: Structure of categories](#)

[Schedule of Events](#) | [Search Abstracts](#) | [Symposia](#) | [Talk Sessions](#) | [Poster Sessions](#)

Aylin Kallmayer<sup>1</sup> ([aylin.kallmayer@gmail.com](mailto:aylin.kallmayer@gmail.com)), Melissa Vo<sup>1</sup>; <sup>1</sup>Goethe University Frankfurt, Scene Grammar Lab, Germany

At its core, vision is the transformation of sensory input into meaningful representations. Understanding the structure of such representational spaces is crucial for understanding efficient visual processing. Evidence suggests that the visual system encodes statistical relationships between objects and their semantic contexts. Recently, however, a more fine-grained framework of hierarchical relations has been formulated (“scene grammar”) according to which scene understanding is driven by real-world object-to-object co-occurrence statistics. More specifically, clusters of frequently co-occurring objects form phrases wherein larger, stationary objects (e.g., sink) anchor predictions towards smaller objects (e.g., toothbrush). Still, we know little about the mechanisms and temporal dynamics of these anchored predictions and whether the processing of individual objects already activates representational spaces characterized by phrasal structures. In the present EEG study, we aimed to quantify shared representations between objects from the same versus a different phrase within the same scene in a MVPA cross-decoding scheme paired with computational modelling to probe the format of shared representations. We presented objects from four different phrases spanning two different scenes (kitchen and bathroom) individually in isolation. Classifiers trained on anchor objects generalized to local objects of the same phrase and reverse, but crucially, not to objects from the same scene, but different phrase. This provides first evidence that phrase-specific object representations are elicited by the perception of individual objects. Computational modelling revealed that high-level semantic features quantified from Resnet50 successfully predicted the classifier’s generalization matrix suggesting that late-stage recurrent processes are responsible for the observed generalization rather than low-level visual similarity between the objects. Overall, we provide novel insights into the temporal dynamics of encoded object co-occurrence statistics which seem to reflect a more fine-tuned hierarchical structure than previously assumed. Finally, this also provides a mechanistic account for the hierarchical predictions observed in efficient attention guidance through real-world scenes.

Acknowledgements: This work was supported by SFB/TRR 26 135 project C7 to Melissa L.-H. Vö and the Hessisches Ministerium für Wissenschaft und Kunst (HMWK; project ‘The Adaptive Mind’) and the Main-Campus-Doctus stipend awarded by the Stiftung Polytechnische Gesellschaft to Aylin Kallmayer.