The neural representation of concepts during composition

Alona Fyshe (afyshe@uvic.ca)
Department of Computer Science, University of Victoria, Victoria, BC V8P 5C2, Canada

Esti Blanco-Elorrieta (eb134@nyu.edu)
Department of Psychology, New York University, New York, NY 10003, USA

Liina Pylkkänen (liina.pylkkanen@nyu.edu)
Department of Psychology, Department of Linguistics, New York University, New York, NY 10003, USA

Abstract

The human brain is able to quickly build complex meaning from simple building blocks. This is especially apparent in language, as we combine words to create phrases, sentences and beyond. Although much research has addressed both the neural representation of individual words and the brain correlates of semantic composition, we do not know how the representations of words evolve and change during composition. Here, we use a picture naming paradigm to explore semantic composition under controlled conditions, wherein participants utter different combinations of adjectives and nouns. We find that, when compared to a non-compositional task, a compositional task 1) has a neural representation that is more similar to the single noun condition, 2) produces a less salient neural representation of the adjective, but 3) produces a more salient representation of the noun. These results are an important first step towards understanding the representation of higher-order meaning in the human brain.

Keywords: MEG; language; semantics; semantic composition

Semantic composition is one of the core processes required for comprehending language. While the neural representation of single words has been characterized to some extent (Mitchell et al., 2008; Sudre et al., 2012), the representation of higher-order meaning has been more elusive. In this study, we detail the changes in semantic representations under conditions that vary the need for composition.

The neural dynamics involved in phrase building have been measured with MEG for both language comprehension (Bemis & Pylkkänen, 2011) and production (Pylkkänen, Bemis, & Blanco-Elorrieta, 2014). Here we explore a new picture naming paradigm that varies compositional requirements. This paradigm is powerful because the stimuli are constant, and the task dictates whether composition is required. Thus, any difference in activity can be attributed only to differences in language processing, not to differences in the stimuli.

In the experiment, participants viewed one of fifty stimuli composed of five object colors crossed with five object types. The background colors were balanced and chosen from the same set of five colors. Each image was presented four times across three conditions. In each of the conditions, participants were asked to describe with a single utterance the identity of the object (noun only condition), the color and identity of the object (phrase condition), or the color of the background and the identity of the object (list condition). For the stimulus in Figure 1a the expected responses for the noun, phrase and list conditions would be lamp, white lamp, and green lamp, respectively. Each trial begins with a fixation cross (300 ms) followed by the picture, which is displayed until timeout (1500 ms) or the onset of speech. We average random sets of 5 trials that share the same expected word utterance to make 4 averaged trials per word and 20 total examples per condition.

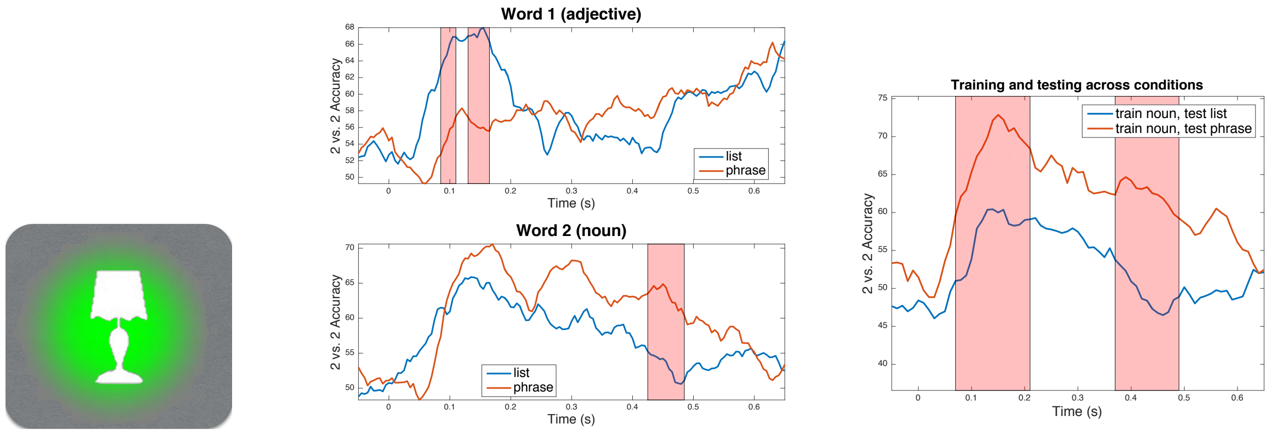
MEG data were collected in the Neuroscience of Language Lab in NYU Abu Dhabi using a whole-head 208 channel axial gradiometer system (Kanazawa Institute of Technology, Kanazawa, Japan). MEG data were recorded at 1000Hz (200Hz low-pass filter), epoched from 100ms before to 600ms after picture onset, and noise was reduced via the Continuously Adjusted Least-Squares Method (Adachi et al., 2001). Artifact rejection was performed as in previous work (Blanco-Elorrieta & Pylkkänen, 2016).

Methods

To represent word semantics, we use pre-trained **word vectors** (Fyshe, Talukdar, Murphy, & Mitchell, 2013). These word vectors are based on the co-occurrence of words in documents across a dataset of millions of webpages. The vectors are compressed using singular value decomposition, and we use the first 100 dimensions for our analysis.

Each trial is associated with three word vectors, one that represents the color of the background ($\mathbf{s^b}$), one that represents the color of the object ($\mathbf{s^a}$), and one that represents the noun ($\mathbf{s^n}$). We train independent ridge regression models on sensor data to predict each dimension of the corresponding word vectors. That is, there are 100 ridge regression models trained for each word type ($\mathbf{s^b}$, $\mathbf{s^a}$, $\mathbf{s^n}$), and the predictions of those models are concatenated to create the predicted background, adjective or noun vectors ($\mathbf{\hat{s}^b}, \mathbf{\hat{s}^a}$ and $\mathbf{\hat{s}^n}$). For each analysis we choose a condition and a word type that is present in the expected response.

To measure performance we use the **2 vs. 2 test** (Mitchell et al., 2008). For each test we selected 2 examples, and trained regression models on the remaining 18. We used the MEG data from the 2 held out phrases to predict 2 word vectors. The 2 vs. 2 task is to choose the correct assignment of predicted vectors $\mathbf{\hat{s}_i}$ and $\mathbf{\hat{s}_j}$ to true vectors $\mathbf{s_i}$ and $\mathbf{s_j}$. We made the choice by choosing the assignment that minimizes the sum of the pairwise cosine distances (d). That is we choose the smaller of $d(\mathbf{s_i}, \mathbf{\hat{s}_i}) + d(\mathbf{s_j}, \mathbf{\hat{s}_j})$ and $d(\mathbf{s_i}, \mathbf{\hat{s}_j}) + d(\mathbf{s_j}, \mathbf{\hat{s}_i})$. The test



(a) Example stimuli: a white lamp on a green background.

(b) 2 vs. 2 accuracy for predicting adjective or noun of expected utterance, in list/phrase conditions. Significant differences are shaded.

(c) 2 vs. 2 accuracy for noun prediction, training on noun only condition, testing on phrase/list conditions.

Figure 1: Example stimuli, and results.

passes if the former (and correct) assignment is chosen; 2 vs. 2 accuracy is the percentage of passed 2 vs. 2 tests. We repeated this process for all pairs of examples where the expected utterances differ (160 in total).

Results

Figure 1b shows results for predicting the adjective or the noun in phrase and list conditions. Each point gives the 2 vs. 2 accuracy when training a model using 100 ms of MEG data centered at the corresponding time. Shaded red areas correspond to significant differences between conditions, determined by cluster permutation tests, $p<0.05$ (Maris & Oostenveld, 2007). We see early differences between conditions when predicting the adjective; the 2 vs. 2 accuracy is higher in the list condition. When predicting the noun, we see high 2 vs. 2 accuracy for both conditions, peaking at about 150 ms, and a significant difference between conditions later in time, when accuracy in the list condition drops.

We were interested to see if neural representations generalized across conditions. In particular, is the noun's neural representation during the noun only condition similar to the representation in the list or phrase conditions? Figure 1c shows the results for this test. Here we train models on the noun only condition, but test on either the list or phrase condition. Training on the noun condition and testing on the phrase condition gives excellent results, peaking slightly higher than training and testing within the phrase condition (72.9% vs. 70.5%). There is a significant difference between testing on the two conditions (windows centered at 70-210 ms and 370-490 ms). This is of particular interest because, again, the stimuli have not changed, only the mental calculations required of the participant. This implies that processing the stimuli to create a noun phrase (i.e. single noun, or adjective noun phrase) is more similar than forming an adjective noun list.

Remaining to be answered are the obvious questions of where composed semantics are represented in the brain, and how brain areas known to be involved in combinatorial processing manipulate those representations. Still, our results are a first step towards understanding the nature of composed representations in the human brain.

Acknowledgments

Alona Fyshe is funded by CIFAR and NSERC. Liina Pylkkänen and Esti Blanco-Elorrieta are funded by National Science Foundation grant BCS-1221723 (LP), grant G1001 from the NYUAD Institute, New York University Abu Dhabi (LP) and La Caixa Foundation fellowship for Post-Graduate Studies (EBE).

References

Bemis, D. K., & Pylkkänen, L. (2011, feb). Simple composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *The Journal of Neuroscience*, *31*(8), 2801–14.

Blanco-Elorrieta, E., & Pylkkänen, L. (2016). Bilingual Language Control in Perception versus Action: MEG Reveals Comprehension Control Mechanisms in Anterior Cingulate Cortex and Domain-General Control of Production in Dorsolateral Prefrontal Cortex. *The Journal of Neuroscience*, *36*(2), 290–301.

Fyshe, A., Talukdar, P., Murphy, B., & Mitchell, T. (2013). Documents and Dependencies: an Exploration of Vector Space Models for Semantic Composition. In *Computational natural language learning.* Sofia, Bulgaria.

Maris, E., & Oostenveld, R. (2007, aug). Nonparametric statistical testing of EEG- and MEG-data. *Journal of neuroscience methods*, *164*(1), 177–90.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., & Just, M. A. (2008, may). Predicting human brain activity associated with the meanings of nouns. *Science (New York, N.Y.)*, *320*(5880), 1191–5.

Pylkkänen, L., Bemis, D. K., & Blanco-Elorrieta, E. (2014). Building phrases in language production: An MEG study of simple composition. *Cognition*, *133*(2), 371–384.

Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R., & Mitchell, T. (2012, may). Tracking Neural Coding of Perceptual and Semantic Features of Concrete Nouns. *NeuroImage*, *62*(1), 463–451.