Representational Similarity Between Brain Activity Elicited by Concrete Nouns and Image Based Semantic Models

Andrew J Anderson,[1] Elia Bruni,[1] Jasper Uijlings,[2] Ulisse Bordignon,[1] Marco Baroni[1] and Massimo Poesio[1]

[1]University of Trento, CIMeC

[2]University of Trento, DISI

How concepts are represented and organised is a key question in cognitive science and is relevant to the design of artificial semantic systems. Neuroimaging techniques allow questions concerning neural representations to be directly addressed. A number of advances have recently been made that ingeniously use models from computational linguistics to predict fMRI activation (Mitchell et al., 2008, Murphy et al., 2009, 2011). In last years V&L NET workshop we suggested that analogous models derived from visual rather than text corpora could serve a similar purpose (Anderson et al., 2011) and we here report results of a pilot analysis.

We re-analyse a subset of Mitchell et al.'s [2008] fMRI data set. During this experiment 9 participants were instructed to actively think about the properties of a concrete noun, when cued with a joint text-image stimulus. Whereas Mitchell et al.'s [2008] original analysis associated neural activity with models derived from stimulus word co-occurrence counts with a set of 25 sensorimotor verbs in a large text corpus (e.g., how often does the word apple co-occur in the same sentence as see, hear, eat, etc). We re-approach the question using visual information.

Recent advances in computer vision have constructed distributional models of meaning, which exploit co-occurrence patterns of words with visual features extracted from tagged images (Bruni et al., 2011). In particular, we adopt three different types of visual semantic representation: **Global**, **Object** and **Surround**. The **Global** representation uses visual information extracted from the whole image. The **Object** representation uses visual information from the segmented object only. The **Surround** representation uses visual information from only outside the segmented object. Interestingly, these visual models have proved to be very effective in capturing meaning when compared to human behavioural measures. In particular, the **Surround** model significantly outperforms the other models, suggesting that the *distributional hypothesis* (i.e., words that are semantically similar appear in similar contexts) transfers also to visual features (see Bruni et al. [2012] for more details).

Representational similarity analyses (Kriegeskorte et al., 2008) were run between fMRI data and the three visual models to test for association in the inter-relationships of categories between modalities. Specifically for each participant, the pairwise similarity between each unique category pairing (45 pairs for the 10 categories for which we had both fMRI data and visual models) was computed using Spearman's rank correlation coefficient (ρ). Following this the 9 per-participant lists of 45 pairwise similarities were collapsed (by averaging) and the resulting list of mean similarities was correlated with the 3 semantic models (again using Spearman's correlation). Significance was tested using a permutation test. Results were: **Global** ρ=0.49, p=0.0004; **Object** ρ=0.278261, p=0.028; **Surround** ρ=0.60, p=0.0002. Consistent with the previous comparison to behavioural measures Surround best explains the brain data.

An implication of these results is that visual context has importance for neural encoding of objects and therefore that the (visual) distributional hypothesis is relevant also for fMRI activation. This is in line with the findings of Mitchell et al. [2008] which highlighted the significance of linguistic context. In future work we shall extend the coverage of artificial concept models and examine how text and image models can be used to segregate neural anatomic regions associated with linguistic and visual experience.

References

Andrew J Anderson, Elia Bruni, Brian Murphy, Marco Baroni, and Massimo Poesio. fmri analyses of semantic structure using joint text and image models. In *V&L NET Workshop on Vision and Language*, 2011.

Elia Bruni, Giang Binh Tran, and Marco Baroni. Distributional semantics from text and images. In *Proceedings of the EMNLP GEMS Workshop*, pages 22–32, Edinburgh, 2011.

Elia Bruni, Jasper Uijlings, Marco Baroni, and Nicu Sebe. Distributional semantics with eyes: Using image analysis to improve computational representations of word meaning. In *Proceedings of ACM Multimedia*, Nara, Japan, 2012. In press.

Nikolaus Kriegeskorte, Marieke Mur, and Peter Bandettini. Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in System Neuroscience*, 2008.

Tom Mitchell, Svetlana Shinkareva, Andrew Carlson, Kai-Min Chang, Vincente Malave, Robert Mason, and Marcel Just. Predicting human brain activity associated with the meanings of nouns. *Science*, 320:1191–1195, 2008.

Brian Murphy, Marco Baroni, and Massimo Poesio. Eeg responds to conceptual stimuli and corpus semantics. In *Proceedings of EMNLP*, 2009.

Brian Murphy, Massimo Poesio, Francesca Bovolo, Lorenzo Bruzzone, Michele Dalponte, and Heba Lakany. Eeg decoding of semantic category reveals distributed representations for single concepts. *Brain and Language*, 117(1):12–22, 2011.