

A perceptual-to-conceptual gradient of word coding along the ventral path

Valentina Borghesani^{1,2,3}, Fabian Pedregosa⁴

1. Université Pierre et Marie Curie, Paris 6, Paris, France,

2. Unicog, INSERM, Gif sur Yvette, France

3. NeuroSpin, DSV, I2BM, CEA, Gif sur Yvette, France

valentina.borghesani@cea.fr

Evelyn Eger^{2,3}, Marco Buiatti^{2,3}, Manuela Piazza^{2,3,5}

4. Parietal, INRIA, Gif sur Yvette, France

5. Center for Mind/Brain Sciences, University of Trento, Italy

Abstract — The application of multivariate approaches to neuroimaging data analysis is providing cognitive neuroscientists with a new perspective on the neural substrate of conceptual knowledge. In this paper we show how the combined use of decoding models and of representational similarity analysis (RSA) can enhance our ability to investigate the inter-categorical distinctions as well as the intra-categorical similarities of neural semantic representations. By means of a linear decoding model, we have been able to predict the category of the words subjects were seeing while undergoing a functional magnetic resonance images (fMRI) acquisition. Moreover, RSA in anatomically defined region of interest (ROIs) revealed a significant correlation with length of words and real item size in primary and secondary visual areas (V1 and V2), while a semantic distance effect was significant in inferotemporal areas (BA37 and BA20). Together, these findings illustrate the possibility to decode the distinctive neural patterns of semantic categories and to investigate the peculiar aspects of the neural representations of each single category. We have in fact been able to show a significant correlation between cognitive and neural semantic distance and to describe the gradient of information coding that characterizes the ventral path: from purely perceptual to purely conceptual. These results would not have been possible without a double exploration of the same dataset by means of decoding models and RSA.

Keywords — fMRI, decoding, representational similarity analysis, semantic representations

I. INTRODUCTION

The neural substrate of semantic memory (i.e. our ability to store and retrieve conceptual representations of the world) is an active field of research in cognitive neuroscience, with many questions yet to be answered [1]. On one hand, clinical neuropsychology studies, above all through the observation of semantic dementia patients' deficits and atrophy, provide evidence that the anterior temporal lobe (ATL) plays a pivotal role in semantic memory [2]. On the other hand, cognitive neuroscience literature has shown that primary and secondary sensory and motor cortices activates while processing words referring to concrete objects, suggesting that conceptual representations might consist in activating the sensory motor attributes of the objects that words refer to [3].

Multivariate pattern analysis (MVPA) of functional magnetic resonance images (fMRI) provides neuroscientist with a powerful tool to investigate neural underpinnings of

mental representations [4]. MVPA tests the presence of a mapping between brain activation patterns and stimuli feature space. The two underlying assumptions shared by all MVPA methods are that information content of neural representations is distributed across populations of neuronal units and that stimuli that are representationally similar will generate similar response patterns across those units. In this framework, so called decoding models can be used to predict a discrete class label (classification problem) or a continuous target (regression problem) based on the pattern of activation that a given stimulus elicits [5]. Likewise, representational similarity analysis (RSA) permits to test whether a given stimuli characteristic is coded by the multivoxel activation of a region [6]. Applying decoding and RSA approaches to the investigation of semantic representations can thus allow, at the same time, to classify items as belonging to one semantic category or the other (e.g., animals vs. tools) and to understand how concepts are organized inside each category.

The hypotheses tested in this paper are (a) that semantic information (i.e. category's membership of the stimuli) can be decoded from brain activity; (b) that different stimuli dimensions (e.g. perceptual vs conceptual features' space) will be encoded in partially different brain regions. One previous study has engaged subjects in a semantic categorization task on pictures of tools and shown, using an RSA approach, that purely perceptual characteristics of pictures (i.e. pixel-wise information) were encoded in primary visual areas while more conceptual dimensions (i.e. the prototypical location and related action) were encoded in more anterior areas of the ventral visual pathway [7]. In the present study we make major step-forward, in that we use words instead of pictures, we engage subjects in an orthogonal task (disentangling semantic operations from semantic representations), and finally we analyze the data with a combination of multiple MVPA techniques.

II. MATERIALS AND METHODS

A. Subjects

11 healthy adult volunteers (3 male) participated in the study. All participants were right-handed (as measured with the Edinburgh handedness questionnaire), had normal or corrected-to-normal vision, and were Italian native speakers. Each participant provided signed informed consent to take part in the study, and all procedures were approved by the local ethical

committee. Participants received a monetary compensation for their participation.

B. Behavioral task and stimuli

The experimental session began with the administration of a behavioral questionnaire in order to gather subject specific cognitive semantic space. The stimuli consisted of 12 names of animals and 12 names of tools selected by a preliminary behavioral experiment that involved 50 Italian native speakers. Previous behavioral and EEG study [8] allowed predicting which words would fall at a close semantic distance, which at the medium distance and which at the far distance. For example, with respect to the word *cow*, the words *sheep*, *camel*, and *whale* are expected to be semantically close, medium, and far, respectively. Subjects were asked to fill in a semantic similarity questionnaire concerning the 24 items selected. Their task was to rate how similar they perceived two given words on a Likert scale from 1 (very different) to 9 (very similar). The questionnaires were constructed as to show each possible combination of two words within the same category. No pair of words from different categories (i.e. one name of an animal and one name of a tool) was shown, to prevent the large difference across categories from overshadowing the smaller, but relevant, differences of items within them. Each the 132 word pair was presented only once (e.g. if *sheep* – *cow* was presented, then *cow* – *sheep* was not). The presentation of the different pairs was randomized between subjects. However, in order to help subjects to focus on small semantic distance within a given category, the 66 pairs concerning tools and the 66 regarding animals were always kept separated, with half of the subjects rating animals before tools and the other half doing the opposite. We verified through statistical analysis that psycholinguistic variables (i.e. words length, number of syllables, number of letter in common, frequency of use, and whether two words had the same initial or ending letter, accent or gender) could not account for the hypothesized semantic distance

C. fMRI task and stimuli

The fMRI experiment used as stimuli the same 12 names of animals and 12 names of tools (hereafter defined as target events) of the previously described behavioral questionnaire. We choose an orthogonal task to ensure at the same time that (1) subjects were paying attention to the stimuli presented and (2) no top-down task-related effect was directly interfering with or generating the observed activations. Subjects were instructed to answer at the presentation of rare odd stimuli (appearing on average 16% of the trials and consisting either in a picture or in a triplet of words referring to one of the targets) whether the odd stimulus was related or not to the previously seen target (1-back task). The hand-answer mapping (“yes” = press left, “no” = press right) was counterbalanced within subjects, and its order was counterbalanced across subjects. Target stimuli (i.e. the 24 words referring to animals or tools) were presented for 1.9 s., at a flickering rate of 1.4 Hz. The inter target interval was jittered of a factor of 0.1 s around the mean value of 1.8 s. The odd events were presented differently according to their modality: images were shown for 2.0 s while definitions appeared as a series of three words each presented for 0.5 s with an interval of 0.2 s between each. The interval

after each odd event was jittered of a factor 0.3 s around the mean value of 2.0 s. Thus, the odd-target onset asynchrony had a minimum of 3.6 s. Within a given fMRI session, the participants underwent 6 runs of 9 min and 40 sec each. Each run contained 4 repetition of each of the 24 targets, 16 odd stimuli, and 24 rest periods (only fixation cross present on screen).

D. MRI and fMRI acquisition

Data were collected at Neurospin (CEA/Saclay, France) with a 3 Tesla Siemens Magnetom TrioTim scanner. Each subject underwent one session that started with 7 minutes of anatomical acquisition and then 6 functional runs. Anatomical images were acquired using a T1 weighted Mprage sagittal scan (voxels size 1x1x1.1mm, 160 slices). Functional images were acquired using an echo-planar imaging (EPI) scan over the whole brain (TR 2.3s; TE 2.3s; FoV 192mm; voxel size 1.5x1.5x1.5mm; 235 repetitions; 82 slices, multi-band acceleration factor 2, GRAPPA 3). The phase acquisition was from posterior to anterior (PA) and the inclination -20° with respect to the subject’s specific AC/PC line.

III. RESULTS

A. Behavioral data analyses

Similarity scores from all subjects were normalized by scaling them between 0 and 1, in order to correct for differences in the ranking scale adopted by the subjects (i.e. for example few subjects never rated any pairs of words as 9). In a second step, normalized data were arranged to create two 12x12 matrices (for animals and tools separately) describing the pairwise semantic distance between words. Next, we computed for both categories the mean distance matrix averaging the distance matrices of all subjects. We then applied multidimensional scaling analysis (MDS, criterion: metric stress, using Matlab Statistics Toolbox¹) to obtain a graphical representation of the cognitive semantic space of our subjects (Figure 1a).

B. fMRI data pre-processing and first level model

Pre-processing of the functional images was run using SPM8² and included slice time correction, realignment of each scan to the first of each given run, co-registration of anatomical and functional images, motion correction, segmentation, and normalization to MNI space. No spatial smoothing was applied. Functional images were then analyzed within the framework of a general linear model. For each run, 35 regressors were included: 24 of interest (12 animals and 12 tools), 4 of no-interest (left and right answers for both images and definitions odd events), 6 motion parameters and 1 constant. Thus, one beta map was estimated for each target events (i.e. words stimuli) for each run. Both subsequent multivariate analyses – decoding and RSA – had

¹ MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.

² Statistical Parametric Mapping, Wellcome Department of Cognitive Neurology, London, UK.

as input data the beta maps corresponding to the 24 target events normalized across conditions separately run by run.

C. Anatomical ROIs definitions

Anatomically defined ROIs (Figure 1b) were built with the SPM toolbox PickAtlas³ proceeding from occipital to temporal along the ventral path (including voxels of both hemispheres): BA17 (13940 voxels), BA18 (69617 voxels), BA19 (65248 voxels), BA37 (65248 voxels), BA20 (28026 voxels), BA38 (27254 voxels).

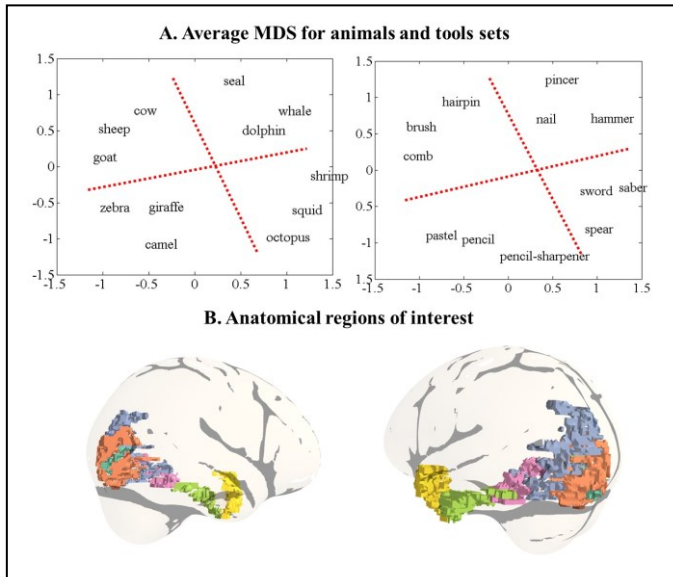


Figure 1 (a) Behavioural results: the cognitive semantic space of the subjects as depicted by the average MDS across their judgments. **(b) Region of interest:** Brodmann areas along the occipital-temporal path selected for the analyses.

D. Decoding

As a first step we investigated the effect of the low level perceptual dimension characterizing the stimuli: the number of letters composing each word. We predict the number of letters (target variable) from the brain activation images in each ROI. This was done with a Ridge regression (regularization parameter = 100) model after selection the most significant 1000 voxels as given by an F-test. We report in Figure 2a the cross-validation scores on 100 folds with a random splitting of 80%-20% between train and test set; the metric used was the coefficient of determination (R2 score). A significant effect of length of words is seen in primary and secondary visual areas.

In a second step, we tested the relationship between activation images and the real size of items (i.e. what we name the “perceptual-conceptual dimension” of our stimuli). In this case, the different stimuli are ordered according to their relative size, so the target variable is of ordinal nature. This can be naturally formulated as a ranking model [9]. The ranking model used was a RankSVM with linear kernel (regularization parameter C=1). We report in Figure 2b the scores (the metric in this case is the mean number of correctly

classified pairs) obtained for each ROIs with the same features selection and cross-validation as above. Real size of the items is significant in all ROIs, peaking in primary and secondary visual areas.

In the third step, in order to decode the category to which the stimuli belonged to (i.e. animals vs tools) we used an SVM model with linear kernel (regularization parameter C=1) on the binary class problem. In Figure 2c we report, for each ROIs, the scores (the metric is again the mean number of correctly classified pairs) obtained with the same features selection and cross-validation as above. The performance is above chance in all ROIs. We also tested the whole brain performance in discriminating between the two classes. Again, we used a linear SVM (C=1) and we applied the same feature selection and cross-validation used for the ROIs analyses. We were able to decode the stimulus category significantly above chance in all subjects.

These results show how the purely perceptual dimension characterizing our stimuli is coded only in visual areas, while more conceptual ones are represented also in more anterior regions. However, in our stimuli there is a covariation between the conceptual-perceptual dimension (real size) and the purely conceptual one (semantic category): indeed, the average real size rank for animals is 17,24 while for tools is 7,23. Hence, it is not possible, with the decoding models here applied, to disentangle the two dimensions. All the analyses of this section were implemented using Scikit-learn⁴.

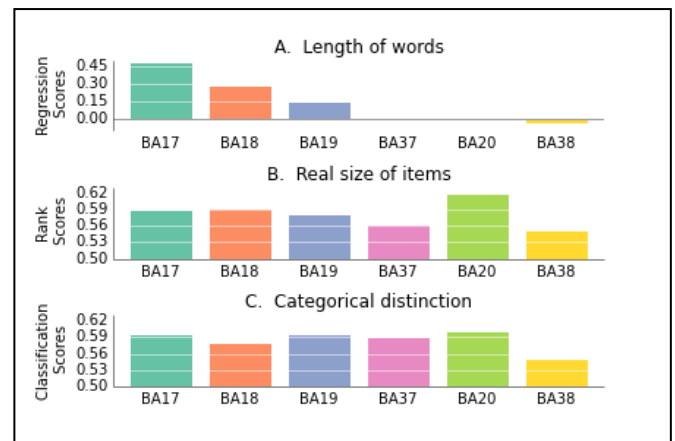


Figure 2 Decoding results: the average effects across subjects of **(a)** length of words, **(b)** real size of items and **(c)** categorical distinction between concepts. For all the metrics, the chance level corresponds to the x axis.

E. Representational similarity analysis

As for decoding, RSA was applied in each ROIs to compare the patterns of neural activity recorded and the predicted matrices illustrating the effect of the different dimensions investigated (Figure 3a). Concerning words’ length the matrix was built computing the pairwise absolute difference: e.g. *sheep* = 5, *cow* = 3, entry *sheep-cow* = 2. The same strategy was applied to real size ranking scale, while the semantic distance matrix was drawn thanks to the behavioural

³ http://www.nitrc.org/projects/wfu_pickatlas

⁴ <http://scikit-learn.org>

data describing the clusters inside each category. To extract the neural matrices, a vector was built with 1000 voxels' value for a given stimulus (i.e. from a given beta map). As for the decoding models, the voxels were chosen with a univariate test, in this case the most active ones in the first level contrast opposing odd stimuli definitions against the implicit baseline. A 24x24 matrix was then generated correlating all the vectors. These matrices were z-transformed and compared (by means of Pearson's correlation) with the predicted ones.

As illustrated by the decoding models, a significant effect of length of words and real item size is observed in primary and secondary visual areas (Figure 3b,c). On the contrary, the purely conceptual dimension (i.e. semantic distance) shows a significant effect only in more anterior areas (Figure 3d). All the analyses of this section were implemented with custom Matlab scripts.

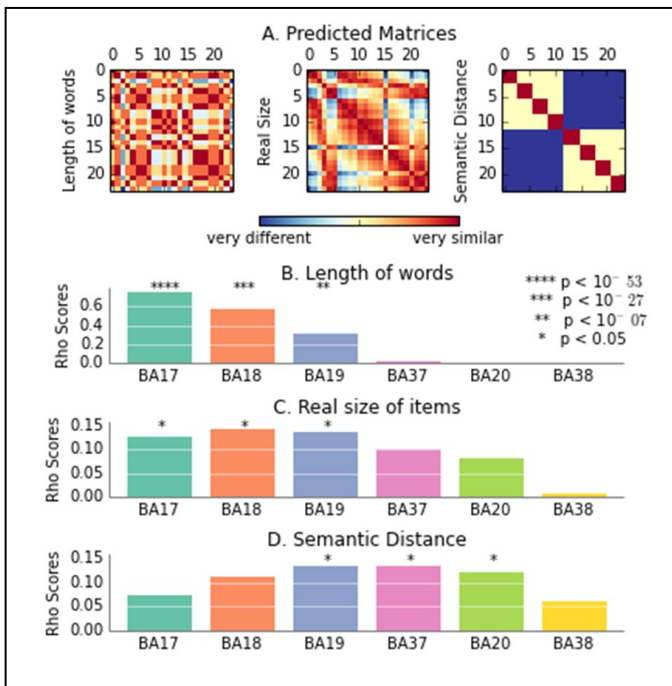


Figure 3 RSA results: (a) predicted matrices illustrating the three dimensions (perceptual, perceptual-conceptual, and conceptual), (b) effects of length of words, (c) real size of items and (d) semantic distance between concepts. Correlations have been tested against zero with a T-test.

The performance of both decoding and RSA models in BA38 can be explained by the drop in signal-to-noise ratio (SNR) observed in this area. For instance, the average SNR in BA18 across subjects was 70.53, in BA38 39.15.

IV. CONCLUSION

In this paper, we have investigated the neural substrate of semantic representations along the ventral visual path. To test our hypotheses, we have developed an fMRI paradigm that (a) engages the subjects in an orthogonal task (i.e. no active semantic categorization was asked), (b) uses as stimuli single words (i.e. a symbolic entry format). Our stimuli (controlled for psycholinguistic variables) varied parametrically along

three different dimensions: a purely perceptual one (i.e. quantified as the number of letters), a perceptual-conceptual one (i.e. the real size of the items referred to by the words), and a purely conceptual one (i.e. the semantic distance characterizing the multidimensional space of concepts the words referred to). These peculiar aspects of our work allowed us to apply a combination of MVPA techniques in order to describe how the different dimensions are represented in the neural pattern of activations of different brain areas along the ventral occipito-temporal path. While a linear SVM was significantly able to discriminate between animals and tools in all the anatomically defined ROIs, RSA results permit to qualify the decoder performance. The correlation between the predicted matrices drawn to reflect the three dimensions and the neural matrices extracted from the 6 ROIs revealed a coding gradient from purely perceptual areas in the occipital lobe to purely conceptual areas in the temporal lobe. Our findings illustrate how neural representations sensitive to different aspects of stimulus characteristics (as physical size on one hand and conceptual field on the other hand) can underlie the differentiation of category-related activation patterns at different levels of the ventral stream hierarchy.

ACKNOWLEDGMENT

Authors would like to thank Bertrand Thirion, Gael Varoquaux and Stanislas Dehaene for fruitful discussions and precious help.

REFERENCES

- [1] Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12), 976-987.
- [2] Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends in cognitive sciences*, 7(8), 354-361.
- [3] Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in cognitive sciences*, 17(9), 458-470.
- [4] Davis, T., & Poldrack, R. A. (2013). Measuring neural representations with fMRI: practices and pitfalls. *Annals of the New York Academy of Sciences*, 1296(1), 108-134.
- [5] Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual review of psychology*, 63, 483-509.
- [6] Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6), 1126-1141.
- [7] Peelen, M. V., & Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *The Journal of Neuroscience*, 32(45), 15728-15736.
- [8] Buiatti M, Finocchiaro C, Caramazza A, Dehaene S, & Piazza M (2012) Word meaning in the human brain: evidence for distinct category specific neural semantic spaces". *Biomag, 18th International Conference on Biomagnetism*, Paris, France.
- [9] Pedregosa, F., Cauvet, E., Varoquaux, G., Pallier, C., Thirion, B., & Gramfort, A. (2012). Learning to rank from medical imaging data. *In Machine Learning in Medical Imaging* (pp. 234-241). Springer Berlin Heidelberg.