NeuroImage 143 (2016) 128–140

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding

Valentina Borghesani [a,b,c,d,*], Fabian Pedregosa [e,f], Marco Buiatti [b,c,d], Alexis Amadon [c], Evelyn Eger [b,c], Manuela Piazza [b,c,d]

[a] École Doctorale Cerveau-Cognition-Comportement, Université Pierre et Marie Curie, Paris 6, 75005 Paris, France
[b] Cognitive Neuroimaging Unit, CEA DRF/I2BM, INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France
[c] NeuroSpin Center, Institute of BioImaging, Commissariat à l'Energie Atomique, F-91191 Gif/Yvette, France
[d] Center for Mind/Brain Sciences, University of Trento, 38068 Rovereto, Italy
[e] Parietal, INRIA, 91191 Gif/Yvette, France
[f] Centre De Recherche en Mathématiques de la Décision, CNRS-UMR, Université PARIS – DAUPHINE, 7534 Paris, France

ARTICLE INFO

Article history:
Received 16 December 2015
Accepted 31 August 2016
Available online 1 September 2016

Keywords:
Semantic knowledge
fMRI
MVPA
Decoding
RSA

ABSTRACT

The meaning of words referring to concrete items is thought of as a multidimensional representation that includes both perceptual (e.g., average size, prototypical color) and conceptual (e.g., taxonomic class) dimensions. Are these different dimensions coded in different brain regions? In healthy human subjects, we tested the presence of a mapping between the implied real object size (a perceptual dimension) and the taxonomic categories at different levels of specificity (conceptual dimensions) of a series of words, and the patterns of brain activity recorded with functional magnetic resonance imaging in six areas along the ventral occipito–temporal cortical path. Combining multivariate pattern classification and representational similarity analysis, we found that the real object size implied by a word appears to be primarily encoded in early visual regions, while the taxonomic category and sub-categorical cluster in more anterior temporal regions. This anteroposterior gradient of information content indicates that different areas along the ventral stream encode complementary dimensions of the semantic space.

Introduction

How is the meaning of words instantiated in the brain? Making sense of symbols involves retrieving from long term memory the *semantic representations* that define what they stand for. One way to think about semantic representations is to consider them as points in a multidimensional space, where each dimension represents a specific property of the concept denoted by the word. In the case of words referring to concrete entities, the semantic space includes both perceptual dimensions (i.e., apprehended through sensory systems; e.g., vision for the prototypical shape, size, or color; audition for prototypical sound) as well as conceptual dimensions (i.e., resulting from a complex combination of multiple sensory-motor ones; e.g., taxonomic class, functional information). Storing both perceptual and conceptual features of object concepts is indeed key for making sense of the word surrounding us, thus for generalizing across conceptually similar but perceptually different objects, and differentiating between perceptually similar but conceptually different ones (Rogers et al., 2004). Consider the words "mouse", "clownfish", "giraffe": thanks to the multidimensional nature of the semantic space we immediately know that the first two refer to animals that are close in size (compared to the third one); that the last two have a similar color (compared to the first one); and that the first and the last one are close in taxonomy (both are terrestrial mammals, compared to the second one, a fish). In this paper, we refer to "perceptual semantic dimensions" of the semantic space as those dimensions along which physical properties of the objects (e.g., size, shape, color, sound) are encoded; and we refer to "conceptual semantic dimensions" of the semantic space as those dimensions along which more complex categorical taxonomic groupings of the objects (e.g., taxonomic class) are encoded.

The question that we approach in this paper is how this multidimensional representational geometry maps onto neural activity. Even though the quest for the neural underpinning of semantics has a longstanding tradition (Martin, 2007; Binder et al., 2009), neither neuropsychology nor functional neuroimaging

* Correspondence to: INSERM-CEA Cognitive Neuroimaging Unit, CEA-Saclay, I2BM, NeuroSpin Bât. 14, 5 - Point Courrier 156, F-91191 Cedex Gif Sur Yvette, France.
E-mail address: valentinaborghesani@gmail.com (V. Borghesani).

http://dx.doi.org/10.1016/j.neuroimage.2016.08.068
1053-8119/

research have provided conclusive evidence on how different perceptual vs. conceptual semantic dimensions defining single concepts are encoded in the brain. Clinical data so far suggest that semantic knowledge is neurally coded in a distributed fashion, as it can be degraded by lesions to sensory–motor brain regions (Pulvermüller and Fadiga, 2010), and profoundly disrupted by lesions to higher–level associative regions (especially the anterior temporal lobe) (Gorno-Tempini et al., 2004; Hodges and Patterson, 2007; Lambon Ralph, 2014). Similarly, functional neuroimaging data indicate that during processing of object-related words there is an increased activation not only in high–level associative cortices (sometimes referred to as "semantic hubs" (Patterson et al., 2007)) such as the inferior frontal cortex (Devlin et al., 2003), the anterior temporal cortex (Mion et al., 2010), or the inferior parietal cortex (Bonner et al., 2013), but also in primary and secondary sensory–motor cortices, in a way that appears proportional to the relevance of perceptuo-motor attributes (Pulvermuller, 2013). Researchers capitalizing from both machine learning techniques and Representational Similarity Analysis (RSA) frameworks have shown that it is possible to discriminate between words belonging to different semantic categories (e.g., animals vs tools) as well as sub-categorical clusters (e.g., mammals vs insects) using distributed patterns of brain activation (Shinkareva et al., 2011; Bruffaerts et al., 2013; Devereux et al., 2013; Fairhall and Caramazza, 2013; Simanova et al., 2014), but they did not determine if such discriminations were driven by conceptual or/and by correlated perceptual information (Naselaris and Kay, 2015). Finally, the so called "encoding" approach (modelling and predicting voxel-wise activation for different stimuli according to their defining set of features) has been successfully applied to predict brain activation during the elaboration of images and movies (Naselaris et al., 2009; Nishimoto et al., 2011), and only very recently to words (Fernandino et al., 2015a). This last study, despite being similar to the present research in that it investigate the semantic coding of symbols (words), only investigated the impact of what we call here "perceptual features" (and not categorical "conceptual" ones). Moreover, it failed to provide a clear picture of the brain topography involved in encoding each of the different features tested. Previous groundbreaking work used a computational model (trained on words data from text corpus) to predict the neural activation associated with written words, but always presented words together with their corresponding picture, thus being unable to dissociate the contribution of low level properties of the physical input from the pure semantic activation driven by the symbolic stimulus (Mitchell et al., 2008).

Here we are interested in studying the types of representations that are evoked by purely symbolic stimuli (written words), and we test the hypothesis that perceptual and conceptual dimensions of the word meaning, for which behavioral studies suggest that they are automatically activated during reading (Rubinsten and

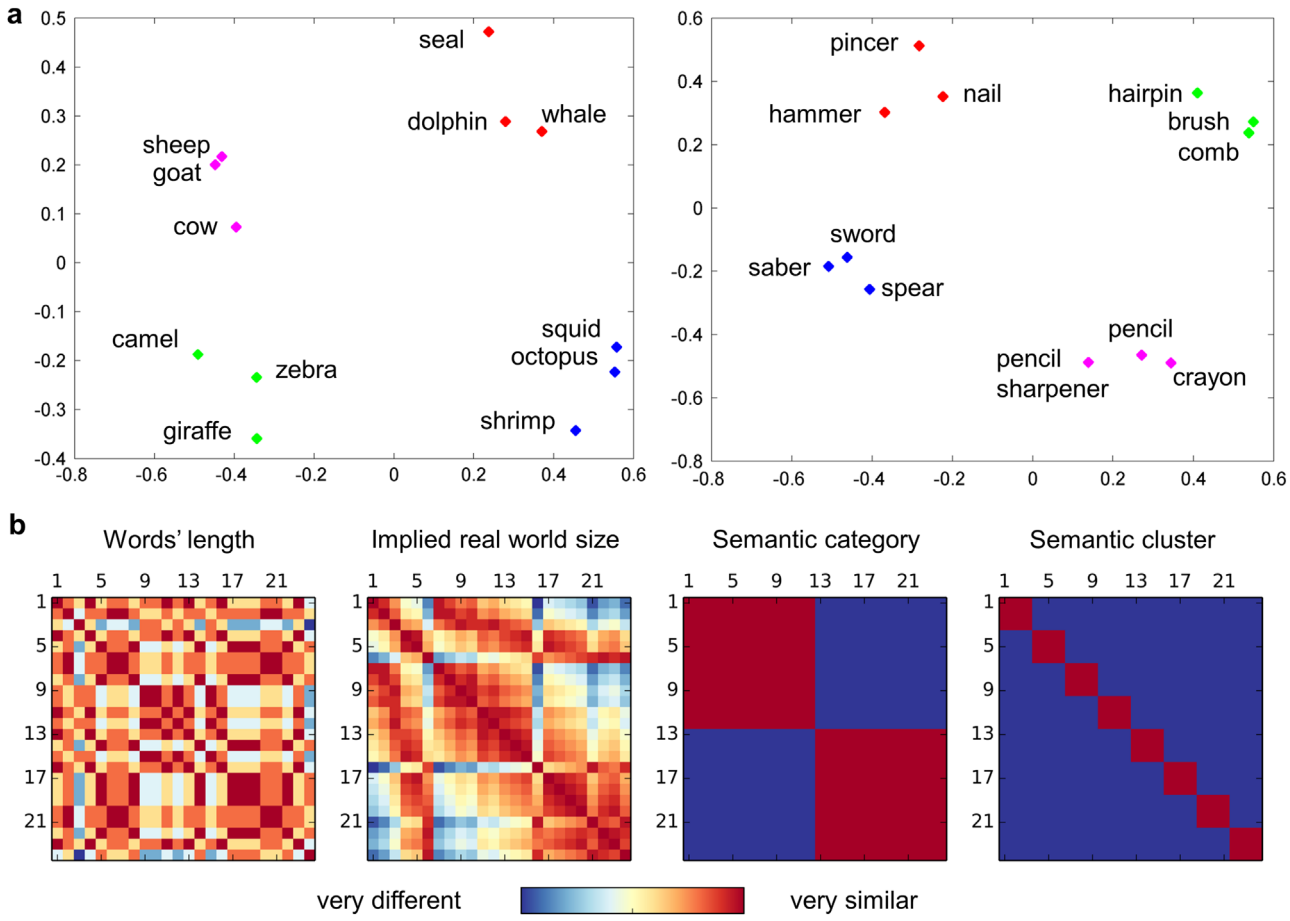

Fig. 1. Words meaning describes a multidimensional semantic space (a) The words used as stimuli in behavioral (a similarity judgment task and a feature generation task) and fMRI experiments. Multidimensional scaling technique was used to visualize the semantic distances perceived between the 12 words denoting animals (left) and the 12 words denoting tools (right). Four clusters of semantically close words are detectable in each of the two semantic categories: domesticated land animals, wild land animals, mammal sea animals, not–mammal sea animals, weapons, office/schools tools, work appliances, and hair instruments. Here shown: the MDS retrieved from the similarity judgment task. (b) Predicted similarity matrices modeling the similarities across stimuli along the four dimensions investigated. The words' length matrix depicts all pairwise differences in terms of number of letters between the stimuli. The implied real–world size matrix is built computing the distances in ranking position between all pairs of stimuli. The semantic category matrix indicates which pairs of stimuli belong to the same category (e.g. both animals) and which do not. The semantic cluster matrix designates which pairs of stimuli belong to the same semantic cluster (e.g. cluster of domesticated land animals: cow, sheep and goat) and which do not.

Henik, 2002; Zwaan et al., 2002; Setti et al., 2009), are coded partially independently in the brain. If that was the case, then we should observe brain regions of which the response profiles reflect dimension–specific metrics, resulting in a double dissociation: some areas should present activation patterns more consistent with the perceptual dimensions of the stimulus space and less with the more conceptual ones (e.g., size, but not taxonomic class), while other areas should present the complementary activation patterns (e.g., more related to taxonomic class and less to size). We presented adult subjects with written words varying parametrically along three different dimensions (Fig. 1a-b): one low level purely physical (the number of letters), one perceptual-semantic (the average real–word size of the objects referred to by the words), and one conceptual-semantic (at two levels of granularity, consisting in 2 semantic categories, each subdivided in 4 sub-categorical clusters). We investigated to what extent the representational geometry of different regions along the ventral visual stream matched the dimension-specific cognitive representational geometry of the stimuli. We predicted that the visual–perceptual semantic dimension of the semantic space would be primarily encoded in early visual regions of the ventral stream (Pulvermuller, 2013), while the conceptual dimensions would be primarily encoded further anteriorly in the temporal lobe (Peelen and Caramazza, 2012). Driven by these predictions, we therefore concentrate our analyses on the ventral visual path.

Materials and methods

Subjects

Sixteen healthy adult volunteers (five males, mean age 30.87 ± 5.34) participated in the fMRI study. All participants were right–handed as measured with the Edinburgh handiness questionnaire, had normal or corrected–to–normal vision, and were Italian native speakers. All experimental procedures were approved by the local ethical committee and each participant provided signed informed consent to take part in the study. Participants received a monetary compensation for their participation. A seventeenth volunteer was excluded from the analyses for not complying with the task (see Testing procedures).

Stimuli

In order to select the target stimuli for the fMRI experiment (i.e. 24 words, 12 names of animals and 12 names of tools) we ran two preliminary behavioral experiments that involved 130 Italian native speakers, tested through internet–based questionnaires.

In the first experiment, we pre-selected 12 animal and 12 tool words, presented fifty subjects with 132 pairs of such words and asked them to rate how similar the concepts indicated by the words were (on a Likert scale from 1 – not similar at all – to 7 – very similar). In order to prevent the large difference across categories from overshadowing the smaller, but relevant, differences within them, we did not pair words belonging to the two different categories and we presented tool word pairs and animal word pairs in separate blocks. The order of presentation of the different pairs inside each category was randomized between subjects while the order of presentation of the two categories was pseudo–randomized: half of the subjects rated animals before tools and the other half did the opposite. All subjects' scores were normalized by scaling them between 0 and 1, in order to correct for possible inter–individual differences in the ranking scale adopted. Normalized data were then re–arranged to create two 12×12 matrices describing the pairwise semantic distance between words for animals and tools separately. Next, for both categories we computed the two mean distance matrices averaging across all subjects. We then applied multidimensional scaling analysis (MDS, 2 dimension, criterion: metric stress) to obtain a graphical representation of the cognitive semantic space of our subjects.

In the second experiment, eighty new subjects took part in a feature generation task: they were asked to list between 5 and 10 characteristics or properties of each of the 24 target stimuli. They were instructed to think about both the physical and perceptual properties (in terms of view, touch, hearing, etc.) and functional properties (e.g. where it is usually found, how and for what it is usually used), as well as any other feature that could be considered important to describe the concepts the word presented referred to. A similarity matrix between the words was then created on the basis of computing how many features were shared by any pair of words belonging to the same category. The subsequent steps (i.e. normalization, conversion in distance matrices and MDS application) were the same as for the similarity judgment task. The goal of collecting these seemingly redundant pieces of data (experiment 1 and 2) was to ensure that the clusters defined were solid (not task dependent) and emerged spontaneously from subjects judgments not only when they were to judge explicitly semantic similarity across word pairs (semantic similarity task) but also when they had to evaluate words individually (features generation task).

Results from the two experiments converge in pointing to 4 sub-categorical clusters in each of the two categories. In the animals set the clusters were: domesticated land animals (cow, sheep, and goat), wild land animals (zebra, camel and giraffe), sea mammals (whale, dolphin and seal), and not–mammal sea animals (squid, shrimp and octopus). In the tools set the clusters were weapons (spear, saber and sword), office/schools tools (pencil, pastel, pencil sharpener), work appliances (hammer, nail, and pincer), and hair instruments (comb, brush, and hairpin) (Fig. 1a).

In order to test the reliability and the consistency of these results, we asked 20 out of the 50 subjects who participated in the similarity judgment experiment to complete the similarity judgment task a second time after 6 months. Subjects received the same instruction as the first time with the added note that it was not a memory task and they should not try to remember the answer given 6 months before. On these data, the same pipeline of analyses described above was applied. The correlation between subjects' similarity matrices was used as a measure of inter–subject variability, while the correlation within subjects was used to estimate the intra–subject consistency. All pairwise across subjects correlations were statistically significant: the average correlation coefficient was 0.73 ± 0.07 for animals and 0.56 ± 0.08 for tools at the first evaluation, and 0.68 ± 0.09 for animals and 0.52 ± 0.05 for tools at the second evaluation. Subjects were also consistent across sessions: all showed a significant and positive correlation between their two judgments for both sets, with an average of 0.78 ± 0.14 for animals and 0.60 ± 0.15 for tools. These analyses were performed with Matlab Statistical Toolbox.

Words belonging to the different semantic categories and clusters were well matched across several psycholinguistic variables such as number of letters, number of syllables, gender, accent and frequency of use (retrieved from *Corpus e Lessico di Frequenza dell'Italiano Scritto* – COLFIS, http://linguistica.sns.it/CoLFIS/Home.htm). These psycholinguistic variables did not differ across the two semantic categories (two–sample *t*-test of frequency: $t=-0.35$, $p=0.73$; number of letters: $t=-1.99$, $p=0.06$; number of syllables: $t=-0.34$, $p=0.74$; chi–square of gender $\chi=0.34$, $p=0.56$; accent $\chi=3.0$, $p=0.08$) or across the four semantic clusters (Kruskal–Wallis test for small sample size of frequency: $h=10.44$, $p=0.17$; number of letters: $h=8.38$, $p=0.30$; number of syllables: $h=9.34$, $p=0.23$; chi–square test of gender: $\chi=6.0$, $p=0.54$; accent: $\chi=2.44$, $p=0.93$). These analyses were run with the

statistical functions provided by Python's library SciPy (http://docs.scipy.org/doc/scipy/reference/stats.html).

Testing procedures

In order to obtain a measure of the subject specific cognitive semantic space and verify the validity of the pre-defined clusters for the subjects participating in our fMRI experiment, we asked our participants to complete the same similarity judgment questionnaire as described above. The experimental session of the main experiment was divided into two parts: first, subjects underwent the fMRI experiment (being totally naïve with respect to the type of stimuli that were going to be presented), then they completed the similarity questionnaire. The analyses of the questionnaires followed the same steps as we used to pre–select the stimuli. To assess the consistency of each subject's judgement with the semantic space that had emerged from our prior behavioral experiments, we computed the correlation between the subject specific normalized distance matrix for animals and tools and the average ones obtained from the fifty subjects that had participated in the first behavioral study. Because one subject failed to comply with the instruction of the task (pressing the response keys according to a numerical progression (1, then 2, then 3, etc…) regardless of the pair of words presented), we excluded his data (both behavioral and fMRI) from further analysis. All sixteen remaining subjects showed positive and significant correlations with the behavioral group average: 0.84 ± 0.08 and 0.84 ± 0.10 for the animals and tools respectively. Because there was very little inter-subject variability in the ratings we decided not to apply a subject specific similarity space in the subsequent fMRI analyses.

During the fMRI experiment, subjects were instructed to silently read the target stimuli (i.e. 12 names of tools and 12 names of animals) and to perform semantic decisions only on extremely rare odd stimuli (Fig. 2a). The odd stimuli appeared on average on 16% of the trials and consisted either in a picture or in a triplet of words referring to one of the targets, promoting both a depictive and a declarative comparison. Subjects pressed a button with the left or the right hand to indicate whether the odd stimulus was related or not to the previously seen target word (1–back task). The hand–answer mapping was counterbalanced within subjects: half of the subjects answered *yes* with the left hand in the first half of the fMRI runs and then *yes* with the right hand in the last half; the other half of the subjects followed the reverse order. The triplets of words defining the target stimuli did not contain any verbs, in order not to stress the functional differences between animals and tools. Such a 1–back oddball task was orthogonal to

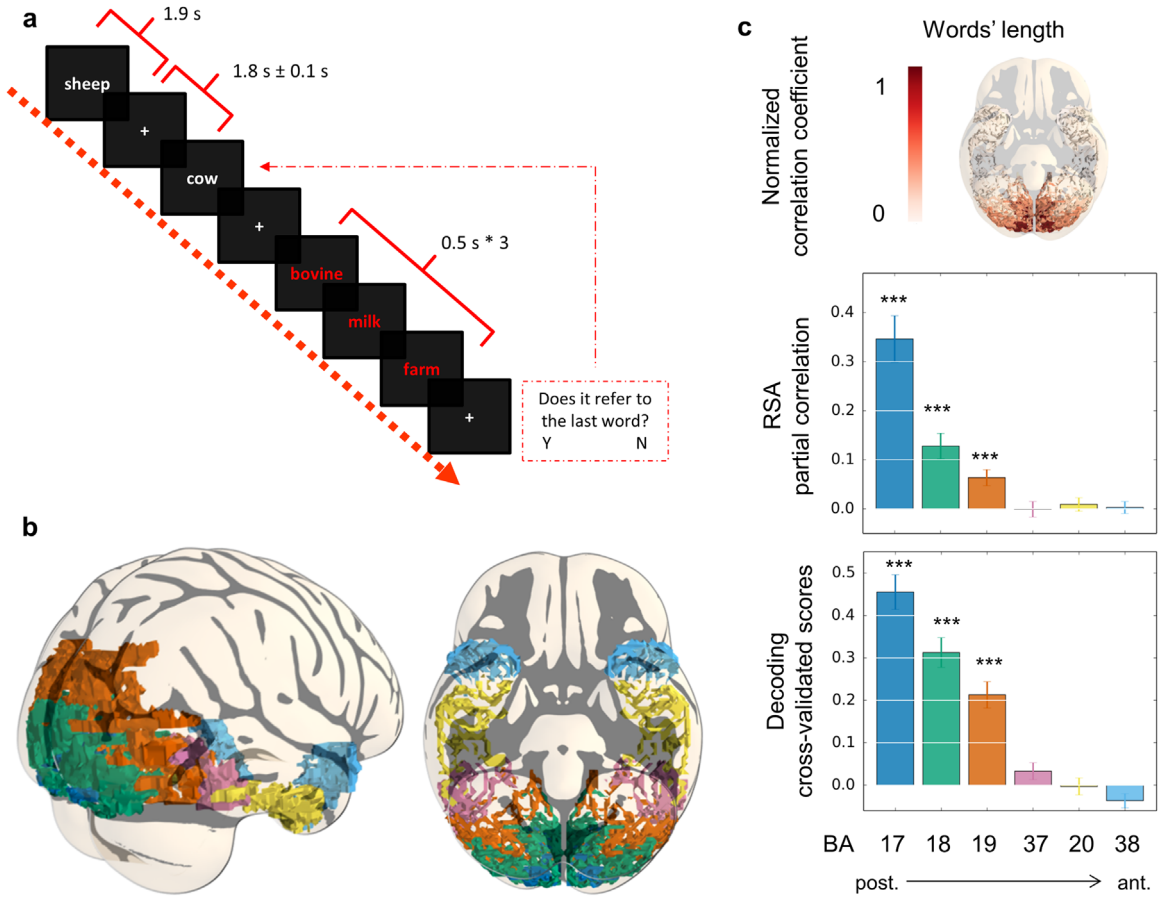

Fig. 2. Experimental setting and low level stimuli representation. (a) Example of a sequence of stimuli: during the fMRI experiment, subjects were instructed to silently read the target stimuli and to press a button at the presentation of rare odd stimuli. The odd stimuli consist either in a picture or in a triplet of words referring to one of the targets. (b) Regions of interest were defined based on anatomical criteria. Proceeding from the occipital lobe to the temporal pole: Brodmann area 17 (primary visual area), Brodmann area 18 (secondary visual areas), Brodmann area 19 (lateral and superior occipital gyri), Brodmann area 37 (occipito–temporal cortex), Brodmann area 20 (inferior temporal gyrus), and Brodmann area 38 (temporal pole). (c) Results concerning the physical dimension of our stimuli (length of the words). Lowermost: the regression model applied (scoring metric: Kendall tau) was able to predict the number of letter composing each word in primary and secondary visual areas. Middle: the partial correlation between neural similarity matrix and length of words matrix is significant in primary and secondary visual areas (while controlling for the other three dimensions investigated). Uppermost: in a template brain, the six ROIs are colored according to the normalized partial correlation scores, highlighting how the effect of the purely physical dimension is confined in occipital visual areas. We are showing the average scores across subjects (no = 16) and error bars indicate the s.e.m. Statistical significance (*$p < 0.05$, **$p < 0.001$, ***$p < 10^{-5}$) is computed with a permutation test and very low p-value are rounded to $p < 10^{-5}$. Exact p-values are reported in the text and **/*** survive Bonferroni correction ($p = 0.05/6$ areas $= 0.0083$).

the dimensions investigated (size, category, cluster), and this allowed us to disentangle task–dependent processes from the spontaneous mental representations of the words (Cukur et al., 2013). Target stimuli were flashed in the center of the screen three times in a row (each time in a different font among Lucida Fax, Helvetica and Courier, to avoid adaptation): each presentation lasted 0.5 s and the interval between them was 0.2 s for a total of 1.9 s for each target stimulus. The goal for this multiple flashed presentation was to ensure that subjects well read the word but at the same time did not have time to make eye movements. The inter target interval was randomly chosen between three values (1.7 s, 1.8 s and 1.9 s, mean = 1.8 s). The odd events were presented differently according to their nature: images were shown for 2.0 s while definitions appeared as a series of three words, presented in a sequence, each for 0.5 s with an interval of 0.2 s between them. The interval after each odd event was randomly chosen between three values (1.7 s, 2 s and 2.3 s, mean = 2 s). The average accuracy in the oddball task was very high = 92.64% (missed = 2.06%, errors = 5.2%). Within a given fMRI session, participants underwent 6 runs of 9 min and 40 sec each. Each run contained 4 repetitions of each of the 24 targets, 16 odd stimuli, and 24 rest periods (only fixation cross present on screen for 1.5 s). Stimuli were completely randomized for each subject and each run, the only constraint being that odd stimuli would appear every 6-to-10 target stimuli. This ensures that, notwithstanding the (minimal) memory component of the task, we can exclude that the results reflect any systematicity due to the stimulus sequence. They were presented with Matlab Psychophysics toolbox (http://psychtoolbox.org/).

MRI protocols

Data were collected at Neurospin (CEA–Inserm/Saclay, France) with a 3 T Siemens Magnetom.

TrioTim scanner using a 32–channel head coil. Each subject underwent one session that started with one anatomical acquisition followed by six functional runs. Anatomical images were acquired using a T_1–weighted MP–RAGE sagittal scan (voxels size 1x1 × 1.1 mm, 160 slices, 7 minutes). Functional images were acquired using an echo–planar imaging (EPI) scan over the whole brain (repetition time = 2.3 s; echo time = 23 ms; field of view = 192 mm; voxel size = 1.5 × 1.5 × 1.5 mm; 235 repetitions; 82 slices, multi–band acceleration factor 2, GRAPPA 3) (Feinberg et al., 2010; Moeller et al., 2010). The acquisition used a phase encoding direction from posterior to anterior (PA) and an inclination of –20° with respect to the subject's specific AC/PC line.

Data pre–processing and first level model

Pre–processing of the raw functional images was conducted with Statistical Parameter Mapping toolbox (SPM8, http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). It included realignment of each scan to the first of each given run, co–registration of anatomical and functional images, segmentation, and normalization to MNI space. No smoothing was applied. For each subject individually, functional images were then analyzed within the framework of a general linear model (GLM). For each of the 6 runs, 35 regressors were included: 24 regressors of interest (corresponding to the onset of the 12 names of animals and 12 names of tools), 4 regressors of no–interest (corresponding to the onset of the odd events – definitions and images – subdivided into those receiving a left hand vs right hand response from the subject), 6 head–motion regressors (i.e. the six–parameter affine transformation estimated during motion correction in the pre–processing) and 1 constant. Fixation baseline was modeled implicitly and regressors were convolved with the standard hemodynamic response function without derivatives. Low–frequency drift terms were removed by a high–pass filter with a cutoff of 128 s. Thus, one beta map was estimated for each target event (i.e. words stimuli) for each run. Both subsequent multivariate analyses – decoding and RSA – had as input data the 24 × 6 beta maps corresponding to the target stimuli normalized across conditions separately run by run (i.e. within each run the values for each given voxel were normalized across conditions to have zero mean and unit variance).

Region of interest

Given our hypothesis and the absence of principled functional localizers, to avoid circularity regions of interests (ROIs) were defined only based on anatomical criteria thanks to SPM toolbox PickAtlas (Fig. 2b). Proceeding from the occipital lobe to the anterior temporal lobe (ATL), we selected six Brodmann areas along the ventral visual pathway: BA 17 – primary visual area, BA 18 – secondary visual areas, BA 19 – lateral and superior occipital gyri, BA 37 – occipito–temporal cortex (includes the posterior fusiform gyrus and the posterior inferior temporal gyrus), BA20 – inferior temporal gyrus, and BA 38 – temporal pole. We included homologue areas from both hemisphere and the average number of voxels of each ROI were: BA17 (13940 voxels), BA18 (69617 voxels), BA19 (65248 voxels), BA37 (65248 voxels), BA20 (28026 voxels), BA38 (27254 voxels). Given the known signal drop out problems in ATL and following previous similar studies (Peelen and Caramazza, 2012), for each subject we calculated the signal–to–fluctuation–noise–ratio (SFNR) map by dividing the mean of the time series (of the first run) by the standard deviation of its residuals once detrended with a second order polynomial (Friedman et al., 2006). This analysis was carried out with the python library nipype (http://nipy.org/nipype). We then computed the average SFNR in each of our ROIs and verified that in all regions this value was above the value of 20 which is usually considered to be the limit for meaningful signal detection (Binder et al., 2011). The average SFNR across the 16 subjects for BA17 was 49.76 ± 5.63, BA18 = 49.34 ± 4.89, BA19 = 52.76 ± 4.7, BA37 = 42.78 ± 3.65, BA20 = 32.87 ± 2.69, and BA38 = 30.99 ± 2.43.

Univariate analyses

For the univariate analyses only, beta maps where smoothed (kernel [4,4,4]). First, two random effects analyses were run searching for regions in which activity was linearly modulated by length of words and implied real world size. Second, random effects analysis was applied to the contrast animals vs tools. Unsurprisingly, the only significant result was a linear effect of length of words in 5 occipital clusters (extent threshold = 100 voxel, $p < 0.001$ FEW corrected) comprising primary and secondary visual cortices. This is in line with the literature on categorical effects on the ventral stream that shows less consistent results when words stimuli are used (as compared with pictures) [for a recent review on the topic: (Bi et al., 2016)].

Multivariate pattern analyses

None of the semantic variables of interest resulted in a dissociation at the univariate analysis level, thus we used multivariate pattern analysis (MVPA) which investigates differences in the distributed patterns of activity over a given cortical region (Davis and Poldrack, 2013). In this framework, the decoding approach aims at predicting one or more classes of stimuli (i.e. “classification problem”) or a continuous target (i.e. “regression problem”) based on the pattern of brain activation elicited by the stimuli. The models are fitted on part of the data (i.e. train set) and tested on left–out data (i.e. test set). Previous studies of semantic

representations used this method to decode the semantic category of words from brain activations patterns, and generalize this categorical discrimination across different input formats (from pictures to words and vice versa) (Shinkareva et al., 2011; Simanova et al., 2014). These studies, however, are limited because: (1) they evaluate the decoding model on the full brain volume, which fails to provide evidence in favor or against the differential contribution of different regions in coding sensory and/or conceptual information (Shinkareva et al., 2011), or (2) they contrast two broad semantic categories (i.e. animals vs tools), without investigating which dimensions of the meaning of the words (i.e. conceptual vs perceptual) drove the observed discriminations (Simanova et al., 2014). A second approach, representational similarity analysis (RSA) (Kriegeskorte et al., 2008), compares the similarity between different stimuli and the one observed between the multivoxel activations patterns elicited by them (i.e. neural similarity). To our knowledge, this approach was deployed only a few times to investigate the processing of symbolic stimuli (words), and no one investigated at the same time the organization of concepts inside and across semantic categories (Bruffaerts et al., 2013; Devereux et al., 2013). Contrary to previous studies, we estimated the similarity of our stimuli considering multiple dimensions at the same time: a low-level physical dimension (number of letters), and three semantic dimensions (a perceptual–semantic: the size of the objects referred to by the words, and two conceptual–semantic dimensions: the category and sub–category cluster). An advantage of RSA is that it permits the investigation of the neural coding of several different dimensions even when those are partially correlated in the stimuli. For example, in the case of our stimuli there was a correlation between semantic category and implied–real world size, in that the implied real world size of the animals was on average larger than that of tools. Using partial correlation as the association metric within RSA (hereafter "partial correlation RSA"), we are robust to the effect of one dimension (e.g. size) while testing for the correlation between the other dimension (e.g. category) and the neural similarity in a given region (Clarke and Tyler, 2014).

Decoding models

We used two different decoding models to solve our four different prediction problems. First, to predict the number of letters composing each word, we applied a regression model in all ROIs. The chosen model was a Ridge regression (linear least squares with l2–norm regularization). The regularization parameter was selected by a nested cross–validation loop. Given the ordinal nature of our problem (i.e. what matters is the rank position, not the absolute value) the metric used to assess the prediction quality was the Kendall rank correlation coefficient (or Kendall tau). The same regression model was used to predict the averaged implied real world size of the objects referred to by the words: all animals and tools where ranked, regardless of their semantic classification, from the smallest (i.e., pencil sharpener) to the biggest (i.e. whale). The ranking scale was devised by the authors considering the average size of the items. When possible, we used information from encyclopedias; when that information was not available, each author gave an approximate estimate and ranked the items independently; it was then verified that the ranks converged [the rank of the items can be found in Supplementary table 1]. Given that in our set of stimuli the object sizes increased logarithmically, the rank, which we used as our size metric, is equivalent to the logarithm of the sizes (correlation between the ranks and the log of the sizes $r2 = 0.98$).

To solve the binary classification problem related with the semantic category (i.e. decode whether a given beta map corresponded to an animal or a tool word) we used a support vector machine (SVM) model with linear kernel. The loss function chosen was squared hinge loss with l2–norm regularization and, again, the regularization parameter was selected by a nested cross–validated loop. Finally, the same model was applied to solve the multiclass problem using a one–vs–rest scheme.

For all decoding models, we report the cross–validation scores computed by averaging the scores of 5 folds with a leave–one–run–out scheme: within each subject data from five out of six runs were used to fit the model and data from the held out run were used to test it. The group–level results were then computed averaging the scores obtained by each subject, and their significance was tested against the empirically estimated random distribution. To obtain such a distribution, the procedure used to obtain the group results was repeated 10,000 times randomly permuting the labels.

The same regression and classification models were fed with the stimuli themselves (i.e., the matrices of 0 and 1 representing the physical appearance of the words used during the experiment, averaging across the three fonts used) to rule out that any of our results could be explained by some low–level characteristic of the stimuli. The goal here is to show that in the stimuli themselves there is already enough information to decode the low level physical dimensions (i.e., number of letters), but not higher level semantic dimensions (nor the perceptual one – size, nor the conceptual one – category and cluster), thus showing that what is retrieved from the patterns of brain activity is not due to any low level property of the stimuli used.

All the analyses described in this section were conducted with the machine learning library in Python Scikit–Learn (http://scikit-learn.org).

RSA

The first step of representational similarity analysis was the modeling of predicted similarity matrices corresponding to the different dimensions investigated. Concerning word length the matrix was built computing the pairwise absolute difference in number of letters between every word pair (the simplest measure of visual similarity). For instance, the entry corresponding to *sheep* (no of letters = 5) vs *cow* (no of letters = 3) would contain a $|5–3| = |2|$. The same strategy was applied to the implied real size ranking scale: the entry corresponding to *whale* (position in ranking = 24) vs *pencil sharpener* (position in ranking = 1) would contain a $|24–1| = |23|$. These first two matrices show distances (i.e. dissimilarity) thus in order to be compared with the neural similarity matrices, their values need to be inverted (similarity = 1–dissimilarity). As to the conceptual dimensions of our stimuli, two matrices were built: one depicting the two semantic categories and one describing the eight clusters that had emerged from the behavioral study. The first one had 1 for all entries of the same category (i.e. all identical combinations: two animals or two tools) and 0 everywhere else (i.e. all different combinations: an animal and a tool). The semantic cluster matrix was built likewise, thus having 1 for all combinations of items from the same cluster and 0 everywhere else. The four matrices being symmetrical (Fig. 1b), they were vectorized discarding the diagonal and keeping only the upper half, then standardized to have mean 0 and standard deviation 1. It should be noted that there is a significant correlation between the similarity matrix of size and the ones of semantic category ($r = 0.39$, $p < 0.001$) and semantic cluster ($r = 0.27$, $p < 0.001$), due to the fact that animal–words tend to refer to big items and tool–words tend to refer to small items. There is, clearly, a correlation between the predicted similarity matrix representing the two semantic categories and the one describing the 8 semantic clusters ($r = 0.32$, $p < 0.001$). Importantly, there is no significant correlation between the predicted similarity matrix for length and the ones for size

(r=0.04, p=0.49), category (r=0.06, p=0.32), or cluster (r=−0.002, p=0.97).

In order to retrieve the neural similarity matrices, for each subject and in each ROI, we built a vector with all the voxels' values for a given stimulus (i.e. from a given beta map). The six stimulus-specific vectors were averaged and all pairwise correlations between vectors were computed (by means of Pearson's correlation). The 24×24 neural similarity matrix obtained was then vectorized as done for the predicted similarity matrices. We obtain thus four vectors (denoted as X_L, X_S, X_C and X_k) from the predicted similarity matrices and one (denoted as Y) from the neural similarity matrix. In order to directly test our hypothesis, we need to be able to estimate the contribution of each single predicted similarity matrix (e.g., X_k) to the neural one (Y) while controlling for the effect of the other ones (e.g., X_L, X_S, X_C). Expressing the neural similarity vector as a linear combination of the predicted similarity vectors plus a noise term, we are interested in testing the null hypothesis that the partial regression coefficient of a given predicted similarity matrix is not significantly different from zero. That is, given the model $Y=\beta1X_L+\beta2X_S+\beta3X_C+\beta4X_K+\varepsilon$ where ε is a vector of residuals, we would like to test the null hypothesis H_0: $\beta_i \neq 0$ (where i can take the values {1, 2, 3, 4}). The test statistic we used for this hypothesis is the partial correlation between all pairs of Y and X (e.g., Y and X_k), controlling for the remaining variables Z (e.g., X_L, X_S, X_C). The partial correlation of two vectors Y and X while controlling for Z is given as the correlation between the residuals RX and RY resulting from the linear regression of X with Z and of Y with Z, respectively. Since the distribution of this statistic is unknown, we choose to obtain the significance level using a permutation test (Anderson and Robinson, 2001). Thus, for each subject and each ROI, we computed the partial correlation between the neural similarity matrices and each predicted similarity matrix (controlling for all the others). The observed result of size is thus corrected for the potential residual correlation between the neural signal and length, category and cluster, the one of category is corrected for length, size and cluster, and so on. Then, scores from all the subjects where averaged and the significance of the group–level results was tested against the empirically estimated random distribution similarly to what has been done for the decoding models. Two features of partial correlation RSA should be noted. First, as Pearson correlation RSA it assumes linear relations between the variables, and the inferences might not be valid if a strong non-linearity underlies the relationship between the physical/cognitive variables and the patterns of brain activation. This issue will need to be tackled in the future to further refine this type of RSA analysis. Second, from a neurobiological point of view, the use of partial RSA can elucidate whether multiple (and partially correlated) features of the stimuli can be independently encoded in the same (set of) brain regions. We think that this question is legitimate, especially in light of the fact that pure functional selectivity (i.e., a brain region in which neurons are solely involved in coding one specific stimulus feature) is clearly not a feature of our brain. It is however necessary to remember that the observation of an interaction between brain region and feature would not imply that a given feature (e.g., size) is solely represented in a given brain region (e.g., visual areas). It would only indicate that there is more residual signal related to a given feature in one area compared to the other. Such results could reflect the fact that more neurons code for one feature in one area than in another one. Alternatively, it may suggest that the different features are encoded with a different degree of precision across areas. The current methods do not allow differentiating across these scenarios: detailed electrophysiological studies might be useful to address the question.

All the analyses described in this section were conducted with in–house python scripts.

Supplementary analyses

We performed five supplementary analyses:

1) In order to demonstrate that our semantic effects (especially those that we could recover from activity in early visual regions) could not be explained by information present in the physical appearance of the stimuli themselves, we applied all the aforementioned decoding and partial correlation RSA analyses to the images of the stimuli (i.e. the snapshots of the screens with the words we presented to the subjects during the fMRI experiment).
2) To better qualify the effect of size as separated, thus independent from the effect of length, even though there was no significant correlation between the predicted similarity matrix for length and size (r=0.04, p=0.49), nor between length and size across the stimuli themselves (r=0.38, p.=0.06), we re-run the partial RSA analyses on a subset of words by removing the two more extreme words length-wise (the shortest and the longest, one animal ("FOCA") and one tool ("TEMPERINO")). This further reduced the already non-significant correlation across Length and Size in our stimuli (down to R=0.27 (p.=0.21)), and the respective distance matrices (down to R=−0.03 (p.=0.5)).
3) To better qualify the presence of different gradients along the ventral stream, we tested for an interaction between the 3 different semantic dimensions (size, category, cluster) and our ROIs by feeding subjects' partial correlation scores (once Fisher r-to-z transformed) into an ANOVA (6 ROIs × 3 dimensions), and then performed trend analyses with SPSS (http://www.ibm.com/analytics/us/en/technology/spss/), testing for a linear, a quadratic, a cubic, a 4-th and a 5-th order term for each of the 3 dimensions.
4) To verify the impact of the partial correlation RSA (vs. standard RSA), we also computed, for all predicted matrices and ROIs, standard Pearson correlation (standard RSA), assessing their significance with permutation tests.
5) Finally, to investigate whether the effects were lateralized, we run an additional partial correlation analysis on the same ROIs but separately for the right and left hemisphere.

Results

In each ROI we applied different MVPA models tailored to our variables and cognitive questions. Firstly, we used decoding to predict: the number of letters composing each word and the relative implied real–size (using the rank from the smallest to the biggest item, approximatively equivalent to the logarithm of the real size), through a regression model; and the conceptual–semantic dimensions at two different scales, that of the semantic category and that of a finer–grained semantic cluster, through a binary classification and a multi–class classification model. We then further qualified the results through partial correlation RSA, and compared the pattern of fMRI activations to words with those predicted by the similarity of the stimulus conditions along the aforementioned dimensions. Extremely low p-value are rounded to $p < 10^{-5}$ and all p-values inferior to 0. 0083 survive Bonferroni correction for multiple ROIs comparisons (p=0.05/6 areas=0.0083).

Physical dimension: number of letters

The number of letters composing each word could be successfully predicted by a regression model in the early visual regions BA17 (mean score=0.45, $p < 10^{-5}$), BA18 (mean score=0.31, $p < 10^{-5}$) and BA19 (mean score=0.21, $p < 10^{-5}$). Likewise, the

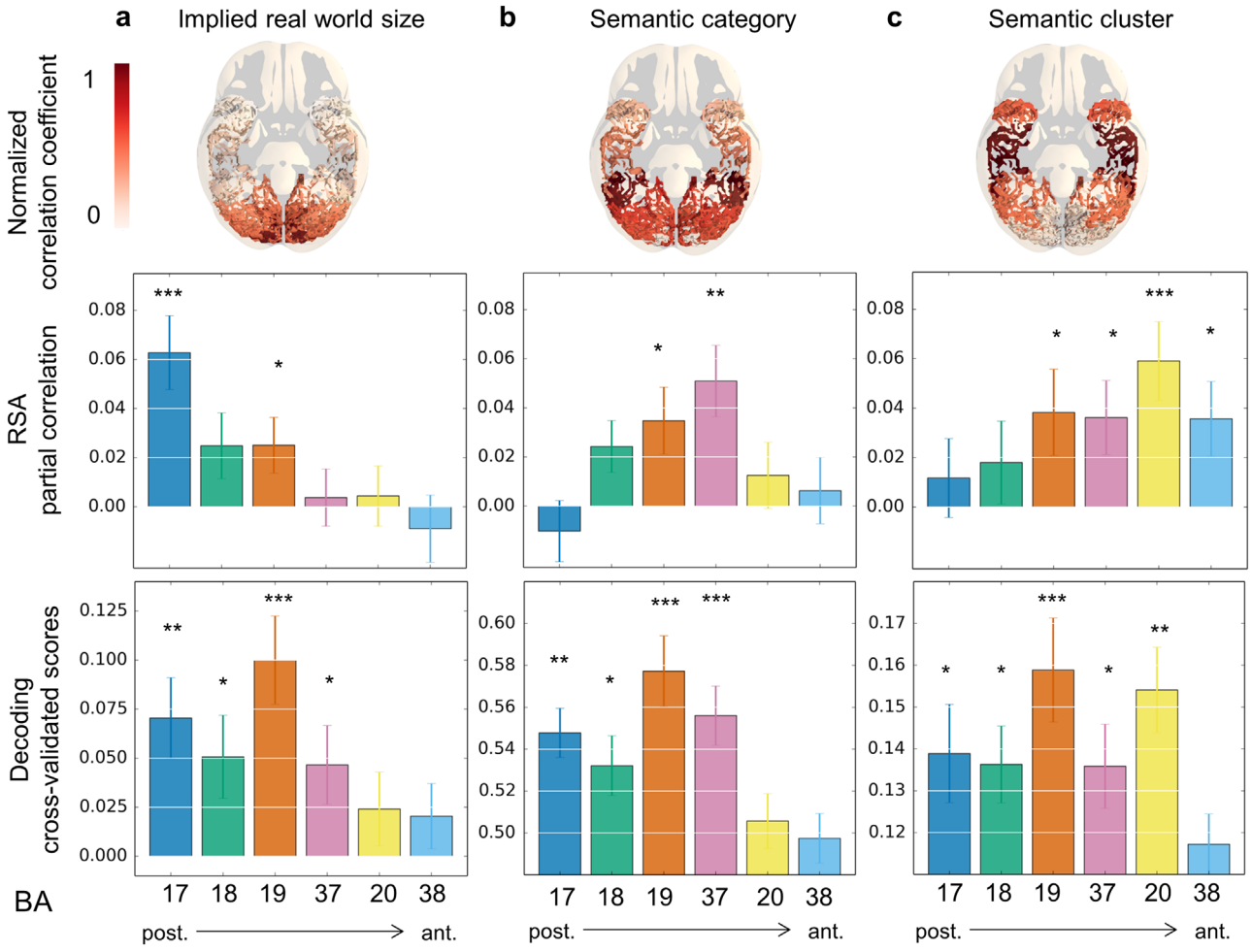

Fig. 3. Topography of perceptual and conceptual representations in the ventral path. (a) Lowermost: the regression model (scoring metric: Kendall tau) was able to predict above chance the implied real-world size in four occipito-temporal areas. Middle: the partial correlation between neural similarity matrix and real–world size matrix, while controlling for the other dimensions, is significant in primary visual areas (BA17). Uppermost: the six ROIs are colored according to the normalized partial correlation scores, highlighting how the effect of the perceptual dimension is confined in occipital visual areas. (b) Lowermost: the binary classification model was able to predict above chance the semantic category in four occipito-temporal areas (from BA17 to BA37). Middle: the partial correlation between neural similarity matrix and semantic category matrix is significant in the occipito–temporal cortex (BA19 and BA37). Uppermost: information about semantic category appears to be coded in occipito-temporal areas, anteriorly respect to the implied real-world size and posteriorly respect to the semantic cluster. (**c**) Lowermost: the multi-classification model was able to predict above chance the semantic cluster in five occipito-temporal areas (from BA17 to BA20). Middle: the partial correlation between neural similarity matrix and semantic cluster matrix is significant in anterior areas, from BA19 to BA38, peaking in BA20. Uppermost: the effect of semantic cluster gets progressively higher the more anterior the areas considered. We are showing the average scores across subjects (no=16) and error bars indicate the s.e.m. Statistical significance ($^{*}p < 0.05$, $^{**}p < 0.001$, $^{***}p < 10^{-5}$) is computed with a permutation test and very low p-value are rounded to $p < 10^{-5}$. Exact p-values are reported in the text and **/*** survive Bonferroni correction (p=0.05/6 areas=0.0083).

neural similarity computed from the pattern of activation of these areas significantly correlated with the predicted similarity matrix modelling the difference in number of letters between each word pair: BA17 (mean score=0.35, $p < 10^{-5}$), BA18 (mean score=0.13, $p < 10^{-5}$) and BA19 (mean score=0.06, $p < 10^{-5}$). More anterior temporal regions ceased to reflect such physical dimension of the visual stimulus, in line with the expected increasing invariance to physical dimensions along the ventral stream. These results are therefore a sound sanity check for our models (Fig. 2c).

Perceptual–semantic dimension: implied real word size

We then investigated the brain code for the real–world size of the objects referred to by the words, to which we refer to as a perceptual–semantic dimension (Fig. 3a). A regression model with the rank of the sizes (equivalent to the log of the sizes) permitted above chance prediction of the relative size in BA17 (mean score=0.07, p=0.0006), BA18 (mean score=0.05, p=0.0086), BA19 (mean score=0.09, $p < 10^{-5}$), and BA37 (mean score=0.04, p=0.0086). Because in our stimuli implied real–word size and semantic category were correlated (on average, tools were smaller than animals) using decoding we were unable to determine if the source of the information used by the decoder to solve the implied real–world size regression problem was indeed related to the implied–real world size, to the semantic category, or both. The partial correlation RSA, on the contrary, could provide such information. Once we accounted for the conceptual effects (semantic category and cluster), the similarity in the implied real–world size significantly correlated with the neural similarity observed in primary visual areas (BA17, mean score=0.06, $p < 10^{-5}$) and then progressively decreased in more anterior areas (BA18, mean score=0.02, p=0.0537, and BA19 mean score=0.03, p=0.0484) (Fig. 3a).

Conceptual–semantic dimensions: semantic category and cluster

Next, we tested more conceptual aspects of our stimuli (Fig. 3b–c): the semantic category (i.e. animals vs tools) and the sub–category semantic clusters (e.g. domesticated animals vs. wild animals). A binary classification model was able to predict above

chance the words' semantic category in four occipito–temporal ROIs: BA17 (mean score=0.54, p=0.0008), BA18 (mean score=0.53, p=0.0055), BA19 (mean score=0.57, $p < 10^{-5}$), BA37 (mean score=0.56, $p < 10^{-5}$). Again, because of the correlation between semantic category and size, these results were further qualified by partial correlation RSA, which showed that category membership was increasingly correlated with brain activation as we moved along the ventral path from posterior to anterior regions (BA18 mean score=0.02, p=0.0558, BA 19 mean score=0.03, p=0.0099), independently from the residual code for size, reaching the peak in BA37 (mean score=0.05, p=0.0004). Finally, using a multiclass classification model we could decode the subtle semantic clustering of our words in five ROIs: BA17 (mean score=0.14, p=0.0126), BA18 (mean score=0.13, p=0.0148), BA19 (mean score=0.16, $p < 10^{-5}$), BA37 (mean score=0.14, p=0.0295), BA20 (mean score=0.15, p=0.0001). These results were further qualified by partial correlation RSA, which showed that semantic cluster membership, once accounted for the other dimensions, was represented in the most anterior areas of the temporal lobe (BA19 mean score=0.04, p=0.006, BA37 mean score=0.03, p=0.0081), peaking in BA20 (mean score=0.06, $p < 0.05$).

Controls on low level physical dimensions

In order to demonstrate that our semantic effects (especially those that we could recover from activity in early visual regions) could not be explained by information present in the physical appearance of the stimuli themselves, we applied all the aforementioned decoding and partial correlation RSA analyses to the images of the stimuli (i.e. the snapshots of the screens with the words we presented to the subjects during the fMRI experiment). Unsurprisingly, the only dimension that this analysis could recover from such input was the number of letters composing each word: decoding score=0.74, $p < 0.001$; RSA score 0.23, $p < 0.001$ (for implied real world size: decoding score=0.12, p=0.28; RSA score=−0.01, p=0.62, for semantic category: decoding score=0.11, p=0.30; RSA score=0.05, p=0.18, for cluster category: decoding score=0.08, p=0.33; RSA score=−0.05, p=0.82).

We also explored if the variations in word length could explain the effect of size in early visual areas. Although the predicted similarity matrices for length and size were not correlated with each other, because the effect of word length was very strong compared to that of size, as a further control aiming at reducing the variability in length across our stimuli we re-run the partial correlation analyses of size eliminating two stimuli, corresponding to the longest (4 letters) and the shortest (9 letters) words. This partial correlation RSA testing for the effect of size (corrected for length, category and cluster) was smaller compared with the one run on the full set of stimuli, but it remained significant ($p. < 0.05$) in BA17. As for the original analysis, this effect disappeared in more anterior regions.

Interaction between semantic dimensions and ROIs

Our findings illustrate two clear postero–anterior gradients in the neural response profile of the ventral visual path: posterior occipital regions appear as coding for the visuo-perceptual semantic property of the items (the implied average real word size), irrespective to their semantic category, while as we moved anteriorly in the ventral stream, mid-anterior temporal regions discriminate first between semantic categories and further anteriorly between sub-categorical cluster in a way that is insensitive to their visuo-perceptual property of size. Such an interaction between semantic dimensions and our ROIs was explicitly tested with an ANOVA (6 ROIs × 3 dimensions). The results was highly significant: $F(10,150)=4.48$, $p < 0.001$, corroborating the differential contribution of perceptual and conceptual semantic dimensions to the pattern of brain activity in occipital and temporal areas. Across the six ROIs, the three effects develop according to different trends: implied real world size shows a significant (decreasing) linear trend ($F(1,15)=23.92$, $p < 0.0001$); semantic category a significant quadratic trend ($F(1,15)=15.97$, p=0.001); semantic cluster a marginal (increasing) linear trend ($F(1,15)=3.59$, p=0.07), not significant likely due to the loss of signal/increased noise in BA38).

Standard pearson correlation RSA

Second, we verified the impact of the use of partial correlation in RSA, and thus run the "standard" Pearson correlation RSA. This revealed a pattern very close to decoding: due to the relation between implied real world size and semantic category/cluster the three effects are intermingled and result in a less clean gradient from physical (length of words: BA17 mean score=0.34, p < 10–5, BA18 mean score=0.12, p < 10–5, BA19 mean score=0.06, p=0.0202) and perceptual (implied real world size: BA17 mean score=0.62, p=0.0133), to conceptual (semantic category: BA37 mean score=0.05, p=0.0446; semantic cluster: BA20 mean score=0.05, p=0.0407) (Sup.Fig. 1).

Lateralization of the effects

Finally, when our ROIs were split in left vs right, the profile of the 4 effects followed the same trend bilaterally: moving from posterior to anterior along the ventral path physical (i.e., length of words) and perceptual (e.g., implied real world size) effects decrease, while conceptual ones (i.e., semantic category and cluster) increase (Sup.Fig. 2). On the left hemisphere, length of words: BA17 mean score=0.33, $p < 10^{-5}$, BA18 mean score=0.14, $p < 10^{-5}$, BA19 mean score=0.05, p=0.0001; implied real world size: BA17 mean score=0.04, p=0.0018, BA19 mean score=0.03, p=0.0102; semantic category: BA18 mean score=0.03, p=0.0112, BA19 mean score=0.03, p=0.012, BA37 mean score=0.06, p < 10–5; semantic cluster: BA19 mean score=0.03, p=0.0048, BA37 mean score=0.04, p=0.0029, BA20 mean score=0.05, p=0.001. On the right hemisphere, length of words: BA17 mean score=0.25, $p < 10^{-5}$, BA18 mean score=0.09, $p < 10^{-5}$, BA19 mean score=0.06, $p < 10^{-5}$; implied real world size: BA17 mean score=0.06, p < 10–5, BA18 mean score=0.02, p=0.0499; semantic category: BA19 mean score=0.03, p=0.0234; semantic cluster: BA19 mean score=0.03, p=0.0134, BA20 mean score=0.05, p=0.0001, BA38 mean score=0.04, p=0.0036. It should be noticed that having now 12 ROIs, the Bonferroni correction threshold is now 0.004 (p=0.05/12 areas=0.004).

Discussion

This study investigated the semantic representation of word meaning along the ventral visual path during silent reading and tested the hypothesis that perceptual semantic features of the objects referred to by the words are encoded in brain regions that are partially segregated from those encoding conceptual semantic features. Our task, orthogonal to the dimensions of the semantic space we investigated, ensured that subjects processed the words at an individual level (as opposed to the category or cluster level), and that the representations recovered in the brain activation emerged spontaneously. Furthermore, since we used words instead of pictures as stimuli, our results are free from any possible low-level confound due to visual shape similarity (Rice et al., 2014). We used a combination of multivariate decoding and partial correlation RSA. In fact, while decoding only tests for the

possibility to discriminate classes (without directly assessing in which aspects those classes differ), partial correlation RSA directly tests for the contribution of a given representational geometry onto brain activity.

Implied real–world size information in primary visual areas

One surprising result of this study is that, during reading, early visual areas appear to contain information relative to at least one perceptual–semantic dimension of word meaning: the implied real–world size of the items they refer to (Fig. 3a). This information, however, is progressively lost towards anterior temporal regions, which become more progressively involved in encoding more abstract information such as semantic category and subcategorical cluster. Not surprisingly, if one had to look only at non-partial correlation RSA or decoding, one would have observed much more distributed effects, with size reaching significance also in more anterior areas and category also in more posterior ones. Having run partial RSA, however, we now know that this would have been a spurious effect due to the correlation between size and semantic category and cluster. Partial correlation RSA gives us a cleaner picture of the contribution of this perceptual dimension once accounting for the conceptual ones. In this respect, it is to be noted that the surprising effect of size in early visual areas was also present when we corrected for the effect of word length, which, even though not significantly correlated with size (neither at the level of the raw values nor at the level of the similarity matrices) was not entirely un-related to it. Further, we could retrieve size-related information in BA17 even after we removed from the analyses the two words that were most greatly variable in length. These results suggest that early visual areas play a role in semantics, and not only in low–level vision, and they are coherent with recent studies indicating that activity in primary visual cortex contains perceptual information even in the absence of sensory stimulation (e.g., the prototypical color of objects presented as a gray-scale image) (Bannert and Bartels, 2013) or in presence of ambiguous stimuli (Vandenbroucke et al., 2014). Our results also relate to the literature on mental imagery, which indicates commonalities between the neural substrates of perception and of imagery (Farah, 1992; Kosslyn, 2001; Smith and Goodale, 2014). In our experiment we neither explicitly prompted the use of mental imagery nor did we inhibit it, thus we are neutral with respect to the issue of whether the observed effects were related to imagery or not. One way to approach the question in the future would be to directly compare the neural representational geometries in early visual cortices during reading (i.e. reading names of objects of different sizes; the condition we have in the present study), with that elicited during perception (i.e. seeing items of different sizes), and mental imagery (i.e. imaging items of different sizes). The recent success of a voxel-wise encoding model suggests that the same low-level visual features are encoded during visual perception and mental imagery (Naselaris et al., 2015); however, further research is needed to test: (1) whether they differ in representational granularity, as is the case for audition and auditory imagery (Linke and Cusack, 2015); and crucially (2) whether similar results are obtained when subjects are presented with symbolic stimuli, i.e. words, instead of pictures. Despite this open issue, however, our results indicate that activation in primary visual areas contains information related to the real–word size of items even when the items are not physically present but simply evoked by symbols. Interestingly, the results of the preliminary behavioral feature generation task we conducted indicate that subjects spontaneously and consistently report size as a key defining property of both animal and tool words (averaging across items and subjects, size-related features were reported 188 times for animals and 212 times for tools), while color, for example, was reported frequently as a feature defining animals but much less for tools (554 times for animals, 117 times for tools). Finally, while the scope of the research was not to investigate the internal scale at which object sizes are represented in the brain, because we computed our dissimilarity matrix on the basis of the rank of the sizes, and because the progression in sizes of our stimuli was roughly logarithmic, our results are compatible with the idea that size is encoded in early visual cortex according to a logarithmic scale (Konkle and Oliva, 2011).

It should be noticed that implied real world size is relatively easily and objectively quantifiable, while other properties, such as color, cannot easily be established for many stimuli. However, in future studies we shall try to parametrize and thus model other visual as well as non-visual sensory properties implied by nouns (e.g., shape, sound) in order to investigate the degree of segregation across sensory regions of these properties. Concerning the anatomy of the real-word size effect, previous literature has shown the implication of lateral-occipital, inferotemporal, and parahippocampal cortices (Konkle and Oliva, 2012; He et al., 2013). The discrepancy between those studies and the current one can be traced down to the numerous methodological differences:

1. Most studies used pictures as stimuli (Konkle and Oliva, 2012 studies 1 and 2), while we used words;
2. When they did not use pictures, but words, as we do, they engaged subjects in tasks involving active size comparison (He et al., 2013) or imagery of objects in their prototypical or atypical size (Konkle and Oliva, 2012 studies 3), thus drawing subject's attention on the size dimension. Instead, in our experiment, subjects were asked to actively think of the whole concept referred to by the words, with no specific focus on the size dimension;
3. Moreover, previous studies compared objects that did not only differ for average size but also belong to largely different semantic categories (animals vs tools vs non-manipulable objects, He et al., 2013), while we present results for the implied real world size effect controlling for categorical differences;
4. Finally, all the aforementioned studies identified the effect of size using univariate analyses, while in our experiment there was no effect, neither in V1 nor in other regions at the univariate level. Multivariate analyses of those data could reveal if additional information could retrieved from brain activity, and especially from primary visual areas, when the distributed pattern of activity is considered.

Conceptual taxonomic information is mainly encoded in mid and anterior temporal areas

A good number of neuropsychological and neuroimaging findings now converge in indicating a crucial role for ATL in the conceptual semantic processing. Herpes simplex encephalitis with widespread lateral and medial temporal lobe damage is associated with semantic category–specific deficits (Lambon Ralph et al., 2007). Moreover, semantic dementia, a neurodegenerative disorder whose gray and white matter atrophy starts in ATL, shows progressive decline in semantic representations spanning all stimulus presentation modalities (visual, auditory, verbal and pictorial) suggesting a key role of ATL in amodal semantic processing. Neuroimaging studies focusing on regions in anterior temporal cortex which are activated during semantic tasks also show that semantic proximity of words belonging to the same semantic category correlates with the patterns of activity in left perirhinal cortex (Bruffaerts et al., 2013). Virtual lesions through TMS and cortical stimulation also indicate that interfering with ATL generates trouble in a variety of semantic tasks (Pobric et al., 2010; Shimotake et al., 2014). These findings are compatible with the

idea that the anterior temporal cortex acts as a hub region where single perceptual sematic features are integrated to give rise to conceptual representations. In the current experiment we show that activity in the mid and anterior temporal cortex (but not in more posterior occipito/temporal regions) reflects categorical and sub-categorical conceptual clustering of the words, and is thus in line with the aforementioned literature. However, because in the current study we investigated at the same time conceptual and perceptual semantic dimensions of the words we presented, we could directly demonstrate that the ATL codes for the conceptual dimensions of the semantic space (category and sub-categorical cluster) in a way that is independent from the single perceptual feature of size. If we had used decoding results only, we would have mistakenly concluded that categorical semantic information is available already in posterior occipital areas. Instead, by partial correlation RSA we can start teasing apart the multiple components of complex representational spaces that characterize word meaning. The finding that even once accounting for the difference across animals and tools in their average size there is enough information in the ATL to discriminate their category and sub-categorical cluster, even if admittedly at a coarse anatomical scale, enriches our understanding of the representational geometry of the anterior part of the temporal lobe. In fact, they complement previous evidence of object category effects in posterior middle/inferior temporal gyrus and ventral temporal cortex (similar to our semantic categories) (Fairhall and Caramazza, 2013), and of semantic similarity effect in left perirhinal cortex (similar to our semantic cluster) (Bruffaerts et al., 2013).

Representational shift along the ventral stream

The third major finding of our study is the observation of two progressive gradients of semantic coding as we move along the ventral stream (Fig. 3): from perceptual to conceptual and from categorical to sub-categorical.

While visuo–perceptual semantic information appears to be preferentially encoded within occipital visual areas, anterior temporal areas become progressively invariant to such perceptual features, and at the same time progressively more sensitive to the conceptual taxonomic dimensions of the semantic space: the semantic category and the sub-categorical cluster of the words. While a similar posterior–to–anterior gradient of abstraction –from physical to perceptual to conceptual information coding– has been previously reported in the domain of object recognition (Peelen and Caramazza, 2012; Devereux et al., 2013; Carlson et al., 2014; Clarke and Tyler, 2014), to our knowledge no study has previously investigated at the same time physical, perceptual and conceptual dimensions of word meaning. The presence of a semantic gradient along the occipito–temporal axis was first suggested by clinical data: patients with vascular damage in the territory of the posterior cerebral artery present fine–grained categorical deficits (e.g. disproportionate failures for biological categories) only if their lesion extend to the anterior temporal region, beyond Talairach's y–coordinate –32 (Capitani et al., 2009). We also observed an increasingly fine-grained clusterization of words as we moved along the anterior temporal lobe: while mid-level temporal regions represent the gross semantic category of the words (animals vs. tools), more anterior regions (BA20 and BA38) become progressively sensitive to the sub-categorical clustering, allowing to distinguish words related, for example, to domesticated land animals, wild land animals, sea mammals, and sea non-mammals. A speculative idea is that the nature of the representation in the temporal lobe could be progressively more fine–grained (i.e. reflecting categorical membership in the posterior portion and single item identity in more anterior one). This hypothesis would also fit well with the report of “concept cells”, coding for individual items though with a very high degree of invariance (even across symbolic and pictorial presentations) in the medial areas of the human anterior temporal cortex (Quiroga, 2012). This representational shift should be interpreted in light of the coarse anatomical scales we used and better qualified by furthers studies tapping the specific representational granularity (or hierarchy) of the different perceptual and conceptual dimensions involved in word meaning in more precisely defined brain regions.

A multidimensional semantic neural space: theoretical implications

Our ROIs encompass several functionally defined areas responding preferentially to different categories of visual stimuli, such as objects (Lerner, 2001), bodies (Downing et al., 2007), faces (Peelen and Downing, 2005) and words (Dehaene and Cohen, 2011). Beside this macroscopic parcellation based on categorical preference, other more abstract dimensions, such as animacy (Sha et al., 2014) and real world size (Konkle and Oliva, 2012) have been suggested as additional organizing principles of object processing in the ventral visual path. In our study we could retrieve size and category information from the activity of occipito–temporal areas, but only at the multivariate level, indicating that the activation of this information during passive word reading is more subtle and distributed compared to that directly evoked by looking at the pictures of the stimuli. Moreover, the discrepancy between findings implicating down-stream regions in the processing of size-related information (Konkle and Oliva, 2012) with our observation of an effect already in early, up-stream, regions could tentatively be explained in terms of differences in task requirements between the two studies (Martin, 2015). Generally speaking, the different perceptual and conceptual dimensions characterizing objects (Huth et al., 2012) and words (Just et al., 2010) semantics appear to be coded in a highly distributed fashion, encompassing visual and nonvisual cortices (Fernandino et al., 2015b). All this evidence contributes to the description of a distributed and multidimensional semantic neural space, partially answering the question of how word meaning is encoded in the brain. A current debate, of interest for some, relates to the question of whether the *format* of the representation of the different stimulus features in the various brain regions is abstract or embodied (Glenberg, 2015; Mahon, 2015). Our study, by investigating the representational geometry of word meaning in different brain regions of the ventral stream elucidates where and how, in the brain, semantic information is encoded. However, it remains neutral as to its *format*. In this respect, we agree with A. Martin (2015) that given the absence of a consensus on how to establish the format of a representation, currently no experimental setting seems to be able to actually tackle this problem. Nevertheless, we think that the double dissociation between coded properties and brain regions that we observed is a convincing argument in favor of a distributed theory of semantic processing that accepts the key role of the anterior temporal lobe in conceptual knowledge and that at the same time recognizes an important part played by sensory-motor areas in encoding perceptual components of meaning.

Conclusion

In conclusion, our results indicate that different aspects of word meaning are encoded in a distributed way across different brain areas. Perceptual semantic aspects, such as the implied real word size appear to be encoded, independently from higher order semantic features, primarily in early sensory areas, which represent the aspects of semantic information that are isomorphic with the input they typically process. Conceptual aspects, such as the categorical cluster and sub-clusters, appear encoded primarily in

anterior temporal areas, which code taxonomic information in a way that is independent from single perceptual features. Hence, both sensory and association areas appear to play an important role by coding for specific and complementary perceptual and conceptual dimensions of the semantic space.

Conflict of interest

The authors declare no competing financial interests.

Acknowledgements

We would like to thank the LBIOM team of the NeuroSpin center for their help in subject scanning, C. Pallier for feedback on the first draft of this manuscript and B. Thirion, G. Varoquaux, and S. Dehaene for fruitful discussions. We gratefully acknowledge K. Ugurbil, E. Yacoub, S. Moeller, E. Auerbach and G. Junqian Xu, from the Center for Magnetic Resonance Research, University of Minnesota, for sharing their pulse sequence and reconstruction algorithms. F.P. benefited from the support of the "Chaire Economie et Gestion des Nouvelles Données", under the auspices of Institut Louis Bachelier, Havas-Media and Paris-Dauphine. The research was funded by INSERM, CEA, Collège de France, and University Paris VI.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuroimage.2016.08.068.

References

Anderson, M.J., Robinson, J., 2001. Permutation tests for linear models. Australian & New Zealand Journal of Statistics 43, 75–88.
Bannert, M.M., Bartels, A., 2013. Decoding the yellow of a gray banana. Curr. Biol. 23, 2268–2272.
Bi, Y., Wang, X., Caramazza, A., 2016. Object domain and modality in the ventral visual pathway. Trends Cogn. Sci. 20, 282–290.
Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex 19, 2767–2796.
Binder, J.R., Gross, W.L., Allendorfer, J.B., Bonilha, L., Chapin, J., Edwards, J.C., Grabowski, T.J., Langfitt, J.T., Loring, D.W., Lowe, M.J., Koenig, K., Morgan, P.S., Ojemann, J.G., Rorden, C., Szaflarski, J.P., Tivarus, M.E., Weaver, K.E., 2011. Mapping anterior temporal lobe language areas with fMRI: a multicenter normative study. NeuroImage 54, 1465–1475.
Bonner, M.F., Peelle, J.E., Cook, P.A., Grossman, M., 2013. Heteromodal conceptual processing in the angular gyrus. NeuroImage 71, 175–186.
Bruffaerts, R., Dupont, P., Peeters, R., De Deyne, S., Storms, G., Vandenberghe, R., 2013. Similarity of fMRI activity patterns in left perirhinal cortex reflects semantic similarity between words. J. Neurosci. 33, 18597–18607.
Capitani, E., Laiacona, M., Pagani, R., Capasso, R., Zampetti, P., Miceli, G., 2009. Posterior cerebral artery infarcts and semantic category dissociations: a study of 28 patients. Brain 132, 965–981.
Carlson, T.A., Simmons, R.A., Kriegeskorte, N., Slevc, L.R., 2014. The emergence of semantic meaning in the ventral temporal pathway. J. Cogn. Neurosci. 26, 120–131.
Clarke, A., Tyler, L.K., 2014. Object-specific semantic coding in human perirhinal cortex. J. Neurosci. 34, 4766–4775.
Cukur, T., Nishimoto, S., Huth, A.G., Gallant, J.L., 2013. Attention during natural vision warps semantic representation across the human brain. Nat. Neurosci. 16, 763–770.
Davis, T., Poldrack, R.A., 2013. Measuring neural representations with fMRI: practices and pitfalls. Ann. N. Y. Acad. Sci. 1296, 108–134.
Dehaene, S., Cohen, L., 2011. The unique role of the visual word form area in reading. Trends Cogn. Sci. 15, 254–262.
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.
Devlin, J., Matthews, P., Rushworth, M., 2003. Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. J. Cogn. Neurosci. 15, 71–84.
Downing, P.E., Wiggett, A.J., Peelen, M.V., 2007. Functional magnetic resonance imaging investigation of overlapping lateral occipitotemporal activations using multi-voxel pattern analysis. J. Neurosci. 27, 226–233.
Fairhall, S.L., Caramazza, A., 2013. Brain regions that represent amodal conceptual knowledge. J. Neurosci. 33, 10552–10558.
Farah, M.J., Soso, Michael J., Dasheiff, Richard M., 1992. Visual angle of the mind's eye before and after unilateral occipital lobectomy. J. Exp. Psychol.: Hum. Percept. Perform., 18.
Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Gunther, M., Glasser, M.F., Miller, K.L., Ugurbil, K., Yacoub, E., 2010. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PloS One 5, e15710.
Fernandino, L., Humphries, C.J., Seidenberg, M.S., Gross, W.L., Conant, L.L., Binder, J. R., 2015a. Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. Neuropsychologia, 76.
Fernandino, L., Binder, J.R., Desai, R.H., Pendl, S.L., Humphries, C.J., Gross, W.L., Conant, L.L., Seidenberg, M.S., 2015b. Concept representation reflects multimodal abstraction: a framework for embodied semantics. Cereb. Cortex.
Friedman, L., Glover, G.H., Fbirn, C., 2006. Reducing interscanner variability of activation in a multicenter fMRI study: controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. NeuroImage 33, 471–481.
Glenberg, A.M., 2015. Few believe the world is flat: how embodiment is changing the scientific understanding of cognition. Can. J. Exp. Psychol. 69, 165–171.
Gorno-Tempini, M.L., Dronkers, N.F., Rankin, K.P., Ogar, J.M., Phengrasamy, L., Rosen, H.J., Johnson, J.K., Weiner, M.W., Miller, B.L., 2004. Cognition and anatomy in three variants of primary progressive aphasia. Ann. Neurol. 55, 335–346.
He, C., Peelen, M.V., Han, Z., Lin, N., Caramazza, A., Bi, Y., 2013. Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. NeuroImage 79, 1–9.
Hodges, J.R., Patterson, K., 2007. Semantic dementia: a unique clinicopathological syndrome. Lancet Neurol. 6, 1004–1014.
Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76, 1210–1224.
Just, M.A., Cherkassky, V.L., Aryal, S., Mitchell, T.M., 2010. A neurosemantic theory of concrete noun representation based on the underlying brain codes. PloS One 5, e8622.
Konkle, T., Oliva, A., 2011. Canonical visual size for real-world objects. J. Exp. Psychol. Hum. Percept. Perform. 37, 23–37.
Konkle, T., Oliva, A., 2012. A real-world size organization of object responses in occipitotemporal cortex. Neuron 74, 1114–1124.
Kosslyn, S.M., Ganis, Giorgio, Thompson, William L., 2001. Neural fundation of imagery. Nat. Rev. Neurosci. 2, 635–642.
Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Bandettini, P.A., 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60, 1126–1141.
Lambon Ralph, M.A., 2014. Neurocognitive insights on conceptual knowledge and its breakdown. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 369, 20120392.
Lambon Ralph, M.A., Lowe, C., Rogers, T.T., 2007. Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model. Brain 130, 1127–1137.
Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., Malach, R., 2001. A hierarchical axis of object processing stages in the human visual cortex. Cereb. Cortex 11, 287–297.
Linke, A.C., Cusack, R., 2015. Flexible information coding in human auditory cortex during perception, imagery, and STM of complex sounds. J. Cogn. Neurosci. 27, 1322–1333.
Mahon, B.Z., 2015. The burden of embodied cognition. Can. J. Exp. Psychol. 69, 172–178.
Martin, A., 2007. The representation of object concepts in the brain. Annu. Rev. Psychol. 58, 25–45.
Martin, A., 2015. GRAPES-Grounding representations in action, perception, and emotion systems: how object properties and categories are represented in the human brain. Psychon. Bull. Rev.
Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y.T., Fryer, T.D., Williams, G.B., Hodges, J.R., Nestor, P.J., 2010. What the left and right anterior fusiform gyri tell us about semantic memory. Brain 133, 3256–3268.
Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A., Just, M.A., 2008. Predicting human brain activity associated with the meanings of nouns. Science 320, 1191–1195.
Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Ugurbil, K., 2010. Multiband multislice GE-EPI at 7 T, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. 63, 1144–1153.
Naselaris, T., Kay, K.N., 2015. Resolving ambiguities of MVPA using explicit models of representation. Trends Cogn. Sci. 19, 551–554.
Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M., Gallant, J.L., 2009. Bayesian reconstruction of natural images from human brain activity. Neuron 63, 902–915.
Naselaris, T., Olman, C.A., Stansbury, D.E., Ugurbil, K., Gallant, J.L., 2015. A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. NeuroImage 105, 215–228.

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. Curr. Biol. 21, 1641–1646.
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. Nat. Rev. Neurosci. 8, 976–987.
Peelen, M.V., Downing, P.E., 2005. Selectivity for the human body in the fusiform gyrus. J. Neurophysiol. 93, 603–608.
Peelen, M.V., Caramazza, A., 2012. Conceptual object representations in human anterior temporal cortex. J. Neurosci. 32, 15728–15736.
Pobric, G., Jefferies, E., Ralph, M.A., 2010. Amodal semantic representations depend on both anterior temporal lobes: evidence from repetitive transcranial magnetic stimulation. Neuropsychologia 48, 1336–1342.
Pulvermuller, F., 2013. How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. Trends Cogn. Sci. 17, 458–470.
Pulvermüller, F., Fadiga, L., 2010. Active perception: sensorimotor circuits as a cortical basis for language. Nat. Rev. Neurosci. 11, 351–360.
Quiroga, R.Q., 2012. Concept cells: the building blocks of declarative memory functions. Nat. Rev. Neurosci. 13, 587–597.
Rice, G.E., Watson, D.M., Hartley, T., Andrews, T.J., 2014. Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. J. Neurosci. 34, 8837–8844.
Rogers, T.T., Lambon Ralph, M.A., Garrard, P., Bozeat, S., McClelland, J.L., Hodges, J.R., Patterson, K., 2004. Structure and deterioration of semantic memory: a neuropsychological and computational investigation. Psychol. Rev. 111, 205–235.
Rubinsten, O., Henik, A., 2002. Is an ant larger than a lion? Acta Psychol. 111, 141–154.
Setti, A., Caramelli, N., Borghi, A.M., 2009. Conceptual information about size of objects in nouns. Eur. J. Cogn. Psychol. 21, 1022–1044.
Sha, L., Haxby, J.V., Abdi, H., Guntupalli, J.S., Oosterhof, N.N., Halchenko, Y.O., Connolly, A.C., 2014. The animacy continuum in the human ventral vision pathway. J. Cogn. Neurosci., 1–14.
Shimotake, A., Matsumoto, R., Ueno, T., Kunieda, T., Saito, S., Hoffman, P., Kikuchi, T., Fukuyama, H., Miyamoto, S., Takahashi, R., Ikeda, A., Lambon Ralph, M.A., 2014. Direct exploration of the role of the ventral anterior temporal lobe in semantic memory: cortical stimulation and local field potential evidence from subdural grid electrodes. Cereb. Cortex.
Shinkareva, S.V., Malave, V.L., Mason, R.A., Mitchell, T.M., Just, M.A., 2011. Commonality of neural representations of words and pictures. NeuroImage 54, 2418–2425.
Simanova, I., Hagoort, P., Oostenveld, R., van Gerven, M.A., 2014. Modality-independent decoding of semantic information from the human brain. Cereb. Cortex 24, 426–434.
Smith, F.W., Goodale, M.A., 2014. Decoding visual object categories in early somatosensory cortex. Cereb. Cortex.
Vandenbroucke, A.R., Fahrenfort, J.J., Meuwese, J.D., Scholte, H.S., Lamme, V.A., 2014. Prior knowledge about objects determines neural color representation in human visual cortex. Cereb. Cortex.
Zwaan, R.A., Stanfield, R.A., Yaxley, R.H., 2002. Language comprehenders mentally represent the shapes of objects. Psychol. Sci. 13, 168–171.