Modality-Independent Decoding of Semantic Information from the Human Brain

Irina Simanova[1], Peter Hagoort[1,2], Robert Oostenveld[1] and Marcel A. J. van Gerven[3]

[1]Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging (DCCN), Radboud University Nijmegen, Nijmegen, The Netherlands, [2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands and [3]Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition (DCC), Radboud University Nijmegen, Nijmegen, The Netherlands

Address correspondence to Irina Simanova, Donders Centre for Cognitive Neuroimaging (DCCN), Radboud University Nijmegen, P.O. Box 9101 NL-6500 HB Nijmegen, The Netherlands. Email: irina.simanova@mpi.nl

An ability to decode semantic information from fMRI spatial patterns has been demonstrated in previous studies mostly for 1 specific input modality. In this study, we aimed to decode semantic category independent of the modality in which an object was presented. Using a searchlight method, we were able to predict the stimulus category from the data while participants performed a semantic categorization task with 4 stimulus modalities (spoken and written names, photographs, and natural sounds). Significant classification performance was achieved in all 4 modalities. Modality-independent decoding was implemented by training and testing the searchlight method across modalities. This allowed the localization of those brain regions, which correctly discriminated between the categories, independent of stimulus modality. The analysis revealed large clusters of voxels in the left inferior temporal cortex and in frontal regions. These voxels also allowed category discrimination in a free recall session where subjects recalled the objects in the absence of external stimuli. The results show that semantic information can be decoded from the fMRI signal independently of the input modality and have clear implications for understanding the functional mechanisms of semantic memory.

Keywords: amodal, FMRI, multivoxel pattern analysis, searchlight, semantics

Introduction

Modern theories about conceptual representation share the view that object properties are stored throughout the brain, with specific perceptual or action-related information stored in corresponding sensory and motor systems. Our knowledge of an object includes its visual features, which will be represented in the brain regions that analyze shape and color, its manner of moving, represented in the brain regions that respond to a particular kind of movement, its taste, involving gustatory areas, and so forth (Riddoch et al. 1988; Martin 2007; Patterson et al. 2007; Mahon and Caramazza 2009). At the same time, the majority of authors agree that these distributed semantic representations can be directly accessed from different kinds of input (Marinkovic et al. 2003; Bright et al. 2004; Price 2004; Kircher et al. 2009; Price 2010; Pobric et al. 2010), i.e. when either a word, a picture, or a characteristic sound refers to an object.

According to modern theories, amodal functionality plays a key role in semantic retrieval (Price 2004). Areas involved in semantic retrieval are taken to represent converging information from multiple modalities and assumed to correspond to high-level object representation. Neuroimaging studies have implicated a number of brain regions as being involved in semantic tasks regardless of input modality. These are ventral

occipitotemporal cortex (Chao et al. 1999; Buckner et al. 2000), posterior middle temporal gyrus (MTG) (Vandenbulcke et al. 2007; Kircher et al. 2009), anterior temporal lobe (Vandenberghe et al. 1996; Patterson et al. 2007), middle frontal gyrus (MFG) (Vandenberghe et al. 1996; Kircher et al. 2009; Van Doren et al. 2010), and left inferior frontal gyrus (IFG) (Bright et al. 2004). As yet, there is a large variation in the effects reported in neuroimaging studies, and there is no definitive anatomical or cognitive model of semantic processing. More knowledge is necessary at the level of neural coding of concepts. That is, what is the common pattern of brain activity across stimulus modalities and how similar are activity patterns in different individuals?

In this study, we addressed the questions stated above by means of multivariate pattern analysis (MVPA) of functional magnetic resonance imaging (fMRI) data. This approach is also referred to as "brain decoding." In 2001, Haxby demonstrated that the category of a viewed object could be identified from the fMRI signal in the ventral visual pathway (Haxby et al. 2001). Since then, extensive work has been carried out on detecting and classifying patterns of activation in response to visual stimuli (Cox and Savoy 2003; Kamitani and Tong 2005; Haynes and Rees 2006; Reddy and Kanwisher 2007; Shinkareva et al. 2008; Shinkareva et al. 2011). In the auditory modality, the ability to decode different speech sounds has also been demonstrated (Formisano et al. 2008; Dasalla et al. 2009). Some studies addressed decoding of the semantic category of visually presented words (Just et al. 2010, Shinkareva et al. 2011), or drawings of objects with captions (Mitchell et al. 2008). Thus, an ability to decode the spatial patterns of fMRI signal associated with object semantics using unimodal means of object presentation has been shown before.

MVPA allows for a direct test of the correspondence between patterns of brain activity when stimuli are presented in different modalities by determining whether a decoding algorithm is able to generalize across modalities. Furthermore, MVPA can be used for mapping patterns of brain activity. A popular approach to pattern localization is to analyze the brain volume with a multivariate searchlight (Kriegeskorte et al. 2006). In this approach, a multivariate statistic is computed in the spherical neighborhood of each voxel as the searchlight sphere moves through the measured volume. The multivariate statistic across spheres allows for detecting non-smooth, but nevertheless functionally localized patterns and thus composes a map of task-related activity in the brain.

Evaluation of the possibility to decode semantic information from different stimulus modalities is also relevant for the development of brain–computer interfaces (BCIs). In BCIs, brain activity is measured and used directly to operate devices such as computers, wheelchairs, or prostheses

(Wolpaw et al. 2002; van Gerven et al. 2009). BCIs are being developed to aid communication and improve the quality of life of people suffering from, e.g., speech impairments. State-of-the-art methods for BCI speech communication involve recording signals from motor cortex involved in speech and translating them into articulation commands that drive a speech synthesizer (Brumberg et al. 2010). Yet, developing a BCI that directly transforms neuronal activity underlying the selection of a particular concept into a speech output is a fascinating challenge for cognitive neuroscience. In this context, identification of the neural signature of modality-independent object representations would be an important stepping stone.

In this study, we applied a support vector machine (SVM) (Vapnik 2000) searchlight technique in a semantic category-decoding paradigm. Subjects performed a semantic categorization task with 2 semantic categories—animals and tools, based on 4 types of stimuli: written names, spoken names, photographs, and natural sounds. In addition to the runs with presented stimuli, there was a free recall run in which subjects were required to recall, or imagine, objects of different categories. We trained the SVM to predict the stimulus category in each of the 4 modalities. We also implemented cross-modal classification by training and testing the SVM across modalities. Both for unimodal and for cross-modal decoding, we analyzed the distribution of SVM prediction accuracies throughout cortex. Finally, we demonstrated that brain regions that allowed modality-independent decoding could predict semantic category in the free recall session.

Materials and Methods

Subjects
Fourteen native Dutch speakers (2 males, 18–27 years of age, 1 left-hander) participated in the study. Data from 3 additional subjects exhibiting head motion beyond 3.0 mm were excluded. All subjects reported that they did not suffer from any neurological disorders and gave written informed consent prior to the experiment in accordance with the guidelines of the local research ethics committee.

Stimuli
Entities of 2 semantic categories (animals and tools) were presented in 4 modalities: verbal auditory (spoken Dutch words recorded digitally at 16 bits with a sampling rate of 44.100 Hz; 20 animal and 20 tool names), verbal visual (written Dutch words, white letters on a black background; 20 animal and 20 tool names), nonverbal visual (colored photographs; 20 animal and 20 tool photographs), and nonverbal auditory (natural sounds; 15 tool sounds and 15 animal vocalizations). The photographs and the natural sounds were obtained from the multimodal stimulus set (Schneider et al. 2008). The exemplar sets in the 4 modalities were not exactly the same: for pictures and natural sounds, the exemplars were selected with special attention to recognition rates and correct category identification, according to the results by Schneider et al. (2008). The spoken and written stimuli were selected such that words belonging to the 2 categories matched for word length. The full list of experimental stimuli is available in the Supplementary Table 1.

Experimental Design
The experimental stimuli in each category in each modality were divided over 2 subsets so that half of the exemplars made up 1 subset (see Supplementary Table 1). Stimuli from 1 subset were presented in 1 block. Within a block, each stimulus was presented 2 or 3 times in succession. Stimulus duration was 400 ms for pictures, written words, and natural sounds. The duration of spoken words ranged from 426 to 1196 ms ($M = 795$ ms, SD = 162 ms). A fixation cross was shown on screen during the auditory presentation. Stimuli were followed by a blank screen with a random duration between 1000 and 1200 ms. The duration of 1 block was 40 s, and the blocks were separated by 10 s of fixation.

Each subset of stimuli was presented 4 times during the experiment but the order of exemplars was randomized upon each next presentation of the block. The order of exemplars was different across all participants.

Subjects were instructed to judge whether each stimulus within a block was semantically consistent with the others. They had to respond upon appearance to out-of-category exemplars. The out-of-category exemplars were distributed randomly in the presentation. In fact, 50% of the blocks did not contain any out-of-category exemplars, 44% contained 1 out-of-category exemplar, and 6% contained 2 such exemplars. All out-of-category exemplars were vehicles, musical instruments, or sport objects. Subjects made responses by pressing a button with the index finger of the dominant hand. The experiment finished with 2 free recall blocks: in each of them, the name of a category appeared on screen for 2 s, followed by 40 s of a blank screen with a fixation cross. Subjects were instructed to recall covertly all the entities from the probed category that they had seen during the experiment.

FMRI Data Acquisition
Subjects were scanned with a Siemens 3T Tim-Trio MRI-scanner, using a 32-channel surface coil. We used a parallel-acquired inhomogeneity-desensitized (PAID) fMRI sequence (Poser et al. 2006). That is, images were acquired at multiple TEs following a single excitation. The following acquisition parameters were used $TE_1 = 9.4$ ms, $TE_2 = 21.2$ ms, $TE_3 = 33$ ms, $TE_4 = 45$ ms; $TR = 2.28$ s. Scans covered the whole brain with 35 slices of 3 mm thickness, slice-gap of 17%; voxel size $3.5 \times 3.5 \times 3.5$ mm^3; field of view 224 mm; echo spacing 0.5 ms. The PAID multiecho sequence entails broadened T_2^* coverage. Because T_2^* mixes into the 4 echoes in a different way, the estimate of T_2^* is improved. A whole-brain high-resolution structural T_1-weighted MPRAGE sequence was performed to characterize subjects' anatomy (TR = 2300 ms, TE = 3.03 ms, 192 slices with voxel size of 1 mm^3, field of view 256 mm), accelerated with GRAPPA parallel imaging.

FMRI Data Preprocessing
The first preprocessing step entailed the combination of images acquired at multiple TEs with a contrast-to-noise ratio weighting approach (Poser et al. 2006) using a homemade MATLAB script (MATLAB R2008b, The MathWorks, Inc., Natic, MA). Subsequent preprocessing of the fMRI data was performed using SPM8 (Welcome Trust Centre for Neuroimaging, University College London, UK). The combined EPI volumes were realigned and spatially normalized to Montreal Neurological Institute space without changing the voxel size. Subsequently, for each subject, a linear model was created, containing the head motion parameters and a set of cosine basis functions. The cosine basis functions captured the low-frequency drifts in the data with a cutoff at 448 s. After model estimation, the resulting residual images were used for further processing.

For the classification, residual images from all volumes collected between 6 s after block onset and 6 s after block offset were selected (correcting for the hemodynamic lag). Images were averaged with a sliding window of 5 images to increase signal-to-noise ratio. A gray matter probability mask was created using a template structural image. Voxels with a gray matter probability exceeding 0.5 were included in the searchlight analysis.

FMRI Data Analysis
Statistical analysis was performed using MATLAB R2008b (The MathWorks, Inc., Natic, MA) and FieldTrip, an open source Matlab toolbox for the analysis of neuroimaging data (Oostenveld et al. 2011). We used a multivariate searchlight technique together with a linear SVM classifier to decode object category (Fig. 1A). The searchlight sphere was centered on each gray matter voxel in turn and the classification
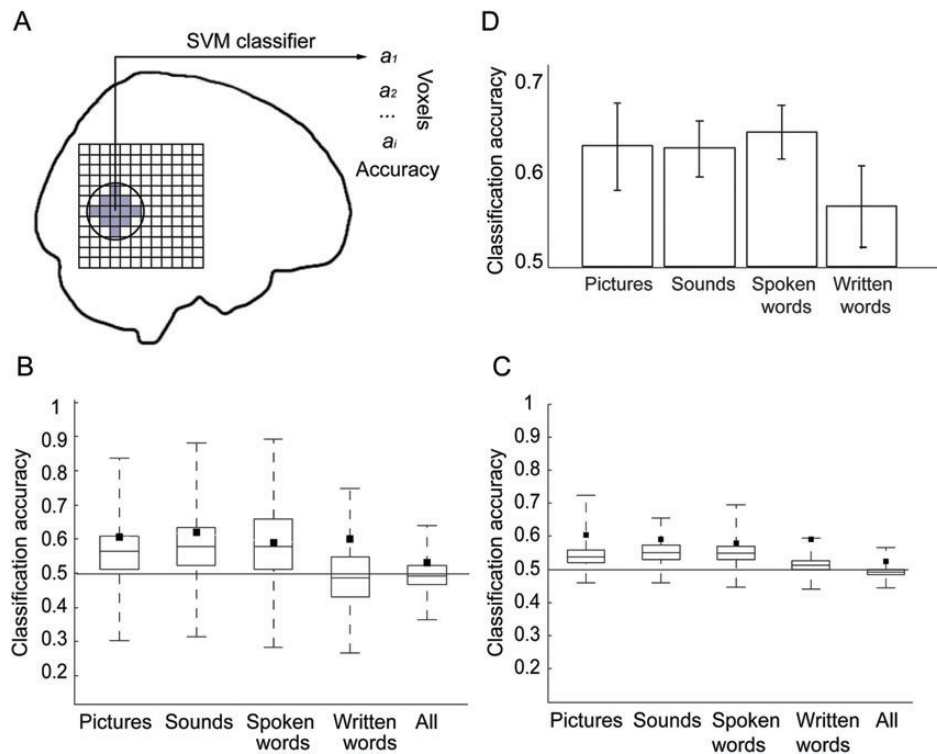
Figure 1. (*A*) Searchlight technique used for decoding object category. The accuracy of the SVM classification was computed in the spherical neighborhood of each voxel as the searchlight sphere moved through the measured volume, probing over 15.000 voxels in total. Resulting accuracy maps were smoothed, thresholded, and used for the group-level analysis. (*B*) Results of the post hoc test showing averaged classification accuracies across 14 subjects for 4 modalities. Selection of voxels for classification was based on the generalization test. (*C*) Distribution of the classification accuracies from the searchlight maps in 1 representative subject. Results of the unimodal and the cross-modal tests are shown. The central mark is the median; whiskers extend to maximal and minimal accuracy values. Black dots indicate the significance threshold (*P* < 0.01 in the permutation test). (*D*) Distribution of the classification accuracies from averaged searchlight maps from 14 participants. Black dots indicate mean significance threshold (*P* < 0.01 in the permutation test) across 14 participants.

accuracy and significance were computed at the local spherical neighborhood of the voxel. The radius of the searchlight sphere was 2.5 voxels (diameter of 17.5 mm) such that each sphere comprised 33 voxels. Classification accuracy (proportion of correctly classified trials) for each sphere was assigned to the sphere's central voxel, in order to produce accuracy maps. The resulting accuracy maps were then smoothed with an 8-mm Gaussian kernel.

The following unimodal classification tests were performed: 1) classification with only picture trials, 2) classification with only natural sound trials, 3) classification with only spoken word trials, and 4) classification with only written word trials. In all unimodal tests, the classifier was trained and tested on the data from different presentation blocks, in which the presented stimuli belonged to different subsets. That is, the testing phase always included unfamiliar category members, which is a crucial ingredient for inferring abstract category representation from MVPA (Vindiola and Wolmetz 2011). The procedure was subsequently repeated with the testing and training subsets interchanged. The results from the 2 validation runs were averaged to produce a single estimate of classification accuracy.

Finally, the generalization test across all modalities was performed. In this test, the classifier was trained on data belonging to 3 input modalities while data belonging to the remaining modality was retained for subsequent validation. The validation procedure was repeated 4 times such that each of the modality-specific data subsets was used once for the validation. The results from the validations were averaged to produce a single estimate of generalization performance.

Significance of the classification outcome at the single-subject level was assessed with a permutation test. The classification accuracies were obtained for 100 samples with random permutations of the class labels. The *P* value was defined as the fraction of samples for which the accuracy was greater than or equal to the accuracy obtained using the correct labeling. While shuffling the labels, we kept track of the

partitioning of the data arising from the cross-validation. That is, the labels were permuted within the partitions of the dataset used to build up each fold (Pereira et al. 2009). Importantly, we also kept track of the original block structure of the experiment. The labels were shuffled not between individual volumes but rather between entire blocks. In that way, we took into account within-block correlations of the fMRI data.

Statistical Group-Level Analysis

Consistency across subjects was accessed with a binomial test. For each searchlight sphere, we calculated the number of significant outcomes (*P* < 0.01 in the permutation test) across 14 subjects. Then, we tested the significance of that observed number with a binomial test under the assumption that the probability of a sphere to be significant in a random subject was 0.01. Thus, we produced an estimate of significant performance of a sphere across subjects under the null hypothesis that the observed number of successful outcomes was due to false positives. In case of rejection of the null hypothesis, activity patterns within a sphere were taken to be informative for the animals–tools classification across participants.

After thresholding, the resulting statistical maps (group-level binomial test, *P* < 0.001, FDR-corrected) were used to identify clusters of interest. The maps were transformed into binary images with all significant voxels assigned to 1. Clusters of significant voxels were identified as the connected components in the 3D binary representation. The resulting clusters were overlaid with the original accuracy maps, and in each of the clusters, the anatomical location of the voxel with maximal classification accuracy (averaged across subjects) was identified. Anatomical labels were derived from the SPM Anatomy Toolbox (Eickhoff et al. 2005).

The left-handed participant was also included in the group analysis. Before including this participant, we ascertained that there was

no significant difference in classification performance between this participant and the others. Furthermore, the patterns emerging in the group-level analysis did not change appreciably, regardless of whether or not this participant was included.

Generalization to the Free Recall Session

An SVM classifier was trained on the data from actual stimulus presentation and then tested on the recall data. Voxels for this analysis were selected individually for each subject. All voxels belonging to the spheres that had shown significant performance ($P < 0.01$) in the generalization across modalities were used to train the classifier. Given the selected voxels, the classifier was trained on the data from each individual modality (4 separate tests) and subsequently tested on the recall session. In addition, the classifier was trained on the combined data from all modalities and tested on the recall session.

Behavioral Analysis

A 2-way analysis of variance was used to test for the differences in the experimental task performance during the blocks of different categories and modalities. This analysis was performed using PASW Statistics 18, Release Version 18.0.0 (SPSS, Inc., 2009, Chicago, IL). This analysis included behavioral data from 13 participants because responses from 1 participant were not recorded due to a hardware error.

Results

Behavioral Results

A significant main effect of modality on the response time was found, $F(3,371) = 17.8$, $P < 0.001$. Participants responded faster to out-of-category items during picture blocks (M = 0.715 s, SD = 0.15) than during written word blocks (M = 0.861 s, SD = 0.15), spoken word blocks (M = 1.026 s, SD = 0.16) and sound blocks (M = 1.160 s, SD = 0.19). When interviewed after the scanning session, all participants reported that the sound blocks represented the most difficult task. This, together with reported response time results, suggests that the task was not of equal difficulty among the modalities. No significant effect of category on response time ($F(1,371) = 0.8$, $P = 0.35$) and no significant interaction (F$(3,371) = 0.402$, $P = 0.751$) was found.

Searchlight Results

The classification accuracies obtained in the whole-brain searchlight analysis are shown in Figure 1. Panel B shows distribution of the accuracies in the searchlight maps of 1 representative subject, and Panel C shows the distribution resulted from averaging the maps from 14 subjects. When averaged voxel-by-voxel across subjects, the accuracy values deviate less from chance level. Extremely low values cancel out when averaging, whereas above-chance values persist in the averaged maps. This indicates that accurate classification was not only possible in individual subjects, but that the same voxels return high classification accuracies across subjects.

For visual object presentations, the group-level statistical analysis of the searchlight maps replicated previous findings, showing categorical differences in regions of the ventral visual pathway (bilateral middle and inferior occipital gyri (IOG), bilateral inferior temporal gyri (ITG), see Fig. 2A and Table 1). Categorical responses in the ventral visual pathway were robust and reproducible across individuals.

For the presentation of natural sounds, the group-level analysis identified category-specific patterns of activation in
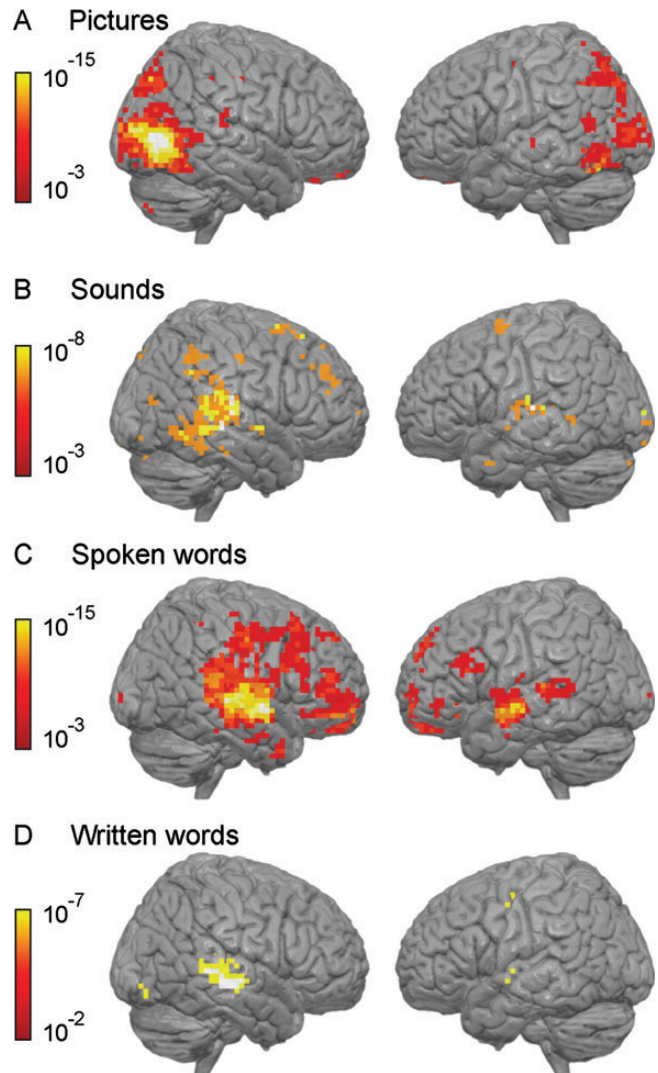

Figure 2. Results of the searchlight analysis, showing discriminability between animals and tools in the within-modality tests. The color represents the *P* values resulting from the group-level statistical analysis. (*A*) Results for picture-based classification. (*B*) Results for natural sound-based classification. (*C*) Results for spoken word-based classification. (*D*) Results for written word-based classification. The maps in panels *A*, *B*, and *C* are thresholded at *P* < 0.001, FDR-corrected; for written words presentations (*D*) the map is thresholded at *P* < 0.01, FDR-corrected, because no clusters survived the more stringent threshold.

primary auditory cortex, specifically, bilateral posterior superior temporal gyri (STG). These patterns can be explained by acoustic differences between the sounds of the 2 categories. Significant clusters in the right hemisphere extend toward the middle temporal and inferior parietal gyri. Additionally, the analysis revealed consistent differential response in bilateral superior frontal gyri (Fig. 2B and Table 1).

In the searchlight classification analysis, both with spoken and written words, a set of voxels in the bilateral STG was identified (Fig. 2C,D), although for written words only few voxels in the left hemisphere survived the threshold. Further, for spoken words, the analysis revealed differentiating voxels in the frontal lobes (right MFG, right SFG, right orbital gyrus, left IFG).

As is evident in Figure 2, in all 4 within-modality tests, group-level statistics revealed larger significant clusters in the

Table 1
The results of the searchlight analysis

Modality	Size	X	Y	Z	Location	P
Pictures	637	48	−70	−1	Right middle temporal gyrus	1.89E−15
	199	−40	−56	−15	Left fusiform gyrus	3.23E−11
	164	−47	−81	−1	Left middle occipital gyrus	9.24E−06
	26	−43	−63	10	Left middle temporal gyrus	9.24E−06
	13	45	−21	45	Right postcentral gyrus	1.86E−07
	12	−22	−95	17	Left middle occipital gyrus	9.24E−06
	11	13	−98	3	Right calcarine	9.24E−06
Sounds	249	62	−46	−1	Right middle temporal gyrus	9.24E−06
	21	−57	−25	10	Left superior temporal gyrus	9.24E−06
	20	−43	−25	6	Left superior temporal gyrus	1.86E−07
	19	55	−49	45	Right inferior parietal gyrus	9.24E−06
	16	27	42	38	Right superior frontal gyrus	9.24E−06
	15	−19	4	69	Left superior frontal gyrus	9.24E−06
	12	24	14	62	Right superior frontal gyrus	9.24E−06
Spoken words	838	66	−14	3	Right superior temporal gyrus	2.85E−13
	386	31	53	−11	Right middle orbital gyrus	1.86E−07
	107	−50	0	−8	Left superior temporal gyrus	3.23E−11
	54	−61	−28	10	Left superior temporal gyrus	9.24E−06
	47	−8	60	24	Left superior medial frontal gyrus	9.24E−06
	39	−43	25	20	Left inferior frontal gyrus	9.24E−06
	23	6	32	45	Right superior medial frontal gyrus	9.24E−06
	15	24	−28	17	Right thalamus	9.24E−06
	14	6	53	41	Right superior medial frontal gyrus	9.24E−06
	13	45	11	−32	Right temporal pole	9.24E−06
	11	−43	42	−1	Left inferior frontal gyrus, p. triangularis	9.24E−06
	11	−1	18	17	Left anterior cingulate	9.24E−06
Written words	114	59	−32	6	Right superior temporal gyrus	9.24E−06
All	301	−43	−53	−11	Left inferior temporal gyrus	3.23E−11
	250	34	39	−8	Right inferior orbital gyrus	3.35E−04
	46	31	−42	−18	Right fusiform gyrus	3.35E−04
	44	−43	−70	27	Left angular gyrus	3.35E−04
	27	31	−70	−50	Right cerebellum	9.24E−06
	27	38	32	3	Right inferior frontal gyrus, p. triangularis	3.35E−04
	17	−40	39	10	Left inferior frontal gyrus	9.24E−06
	12	3	−70	−18	Right cerebellum, vermis	3.35E−04
	11	−40	7	10	Left inferior frontal gyrus, p. opercularis	9.24E−06

After thresholding the group-level statistical maps, clusters of significant voxels were identified. A threshold of $P < 0.001$ was chosen in all cases except for the written words and for the generalization across modalities. For the written words and for the generalization, a threshold of $P < 0.01$ was chosen since, for those modalities, no clusters survived the more stringent threshold. All thresholds were FDR-corrected. In each of the resulting clusters, the voxel with maximal classification accuracy (averaged across subjects) was identified. For each cluster, the table shows the cluster size as well as the anatomical location and label of the identified voxel. For this voxel, the P value obtained in the group-level test is given. Only clusters larger than 10 voxels are listed in the table.

right rather than in the left cerebral hemisphere. One might conclude that the right hemisphere contributes more strongly to categorization task than the left hemisphere across all modalities. However, inspection of within-modality-activation maps from individual subjects showed no prevalence of the right hemisphere in the number of significant voxels. Most likely, the hemispherical asymmetry emerges in the group-level analysis because the category-responsive patterns were more spatially focused and more consistent across participants in the right hemisphere. Therefore, more voxels could have passed the group-level statistical test.

Figure 3 shows the statistical map obtained in the generalization test. The localization results are listed at the bottom of Table 1. The cross-modal searchlight mapping revealed a large group of voxels with maximal accuracy in left ITG. The cluster partly extends to a number of adjacent brain areas, such as left FG, left ITG, left MTG, and left IOG. The second-largest cluster of voxels was identified frontally, covering a large portion of the right IFG and MFG. Further, 2 significant clusters of voxels were identified in the left IFG. Finally, smaller clusters were found in the right FG, left angular gyrus, and right cerebellum.

The analysis did not show a one-to-one correspondence between the areas involved in unimodal and cross-modal category discrimination. This is not surprising given that the statistical sensitivity of the unimodal and cross-modal tests was

different. Figure 1B illustrates the effect of the different thresholds applied in the cross-modal and unimodal tests, showing that the classifiers' predictions in the cross-modal test were less variable than in the unimodal tests. This is due to the fact that the cross-modal test used many more exemplars for both training and testing. With an increase in the number of test trials, the significance threshold decreases because, given a large number of trials, even classification performance that is slightly above chance level does not occur often in a permutation sample.

To further test the functional involvement of the identified cross-modal areas, we performed a post hoc analysis, where an SVM classifier was trained and tested for each of the 4 modalities, using the voxels that had been identified in the cross-modal test (Fig. 1D). As in the whole-brain within-modality analysis, the classifier was always tested on unfamiliar category exemplars. The classifier indeed performed above chance level in each of the 4 modalities, showing that these regions were involved in unimodal category discrimination. The performance is, however, not high enough to pass the threshold in the unimodal tests.

Free Recall Test

As an additional task, at the very end of the experiment, subjects were asked to recall all the exemplars of both categories
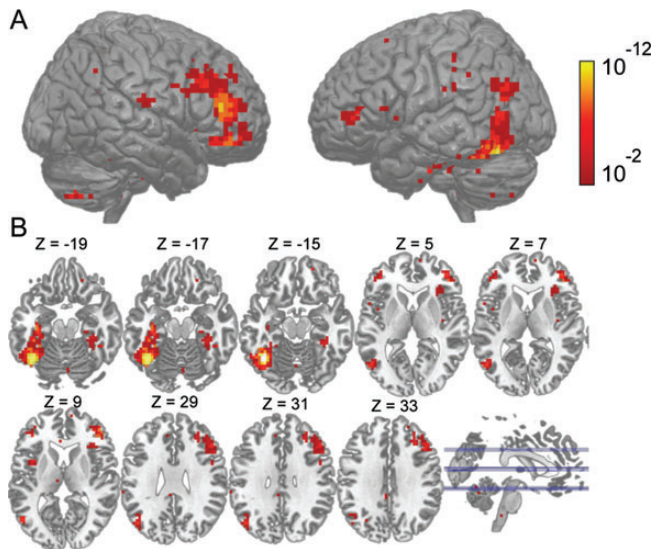
Figure 3. Results of the searchlight analysis, showing discriminability between animals and tools in the generalization across modalities. The color represents the *P* values resulting from the group-level statistical analysis, thresholded at *P* < 0.01, FDR-corrected. (*A*) Whole-brain view. (*B*) Multiple slices view. Locations of the axial sections are illustrated in the bottom row.

Figure 4. Free recall test performance. The averaged classification accuracies across 14 subjects are shown. Selection of voxels for classification was based on the generalization test. The classifiers have been trained on the data from 1 modality, or on the combination from all modalities (see captions below the graph) and tested on the free recall data. Error bars indicate standard error of the mean. *P* values obtained in a right-tail *t*-test against 0.5 are shown above each bar.

they had seen during the experiment. We assessed how well the classifier that had been trained on the data from each individual modality or on the combination of modalities would predict the category of entities in the recall session. Results of this test are shown in Figure 4. Voxels that had shown significant performance in the generalization were used for the analysis in all the modalities. The average classification accuracies and the standard error of the mean in this test were 0.56 ± 0.06, 0.65 ± 0.06, 0.56 ± 0.06, and 0.60 ± 0.07 for classifiers trained on pictures, sounds, spoken words, or written words, respectively. Finally, when trained on the combination of all modalities, and tested on the free recall data, the classifier performed at 0.67 ± 0.05.

In this test, we observed a large variation in classification performance across subjects. One cause of this variation is the low number of exemplars in the test set in this analysis. The recall session only consisted of 2 blocks of 40 s each. Therefore, the number of volumes available for testing the classifier in this condition was relatively low. Classification accuracies deviated considerably from chance level, and even very high accuracies obtained in some participants were not statistically reliable. Nevertheless, Figure 4 illustrates, that, on the whole, across the group of 14 subjects, the free recall prediction accuracies in most of the tests were above chance level. The accuracies obtained from the group were significantly higher than chance level (right-tailed *t*-test, *P* < 0.01), when trained on the combination of the data from all modalities. This indicates successful object category decoding during free recall.

Discussion

In the present study, we investigated the possibility to decode semantic category (animals vs. tools) from fMRI spatial patterns using an SVM searchlight technique in a setting where stimuli were presented using 4 different input modalities. Significant decodin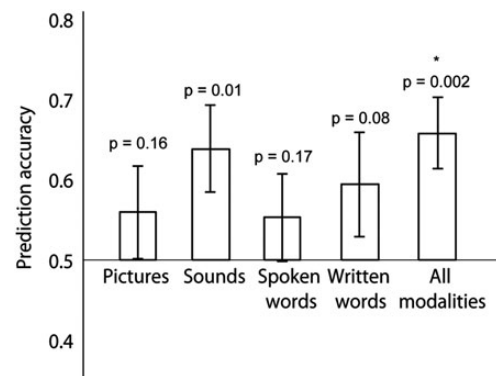g performance was achieved, both within and across different input modalities. Furthermore, the employed searchlight technique allowed localization of those brain regions that were involved in successful decoding across subjects. A free recall test showed that these brain regions even allowed decoding of semantic category in some subjects in the absence of external stimuli.

Overall, the unimodal results in the present study confirm previous work on animal–tool dissociations in the brain. Evidence has accumulated that visual presentation of these object categories elicit distinct patterns of neural activity in regions that mediate object recognition, notably the ventral occipitotemporal processing stream (Martin 2007; Op de Beeck et al. 2008). In a series of studies, Chao et al. (1999, 2002) found greater activity in the lateral fusiform gyrus (FG) for animals than for tools stimuli. In contrast, the medial FG was more active for tools than for animals. In posterior lateral temporal cortex, enhanced activity in left posterior MTG and inferior temporal sulcus in response to tools was reported. These results have been replicated in a number of later studies (reviewed by Martin and Chao 2001; Martin 2007). Importantly, recent studies have reported similar animal-tool distinctions in semantic decision tasks with verbal stimuli, namely written objects' names (Devlin et al. 2005; Hauk et al. 2008) and spoken objects' names (Noppeney et al. 2006). Only few studies have investigated dissociations in processing of animal and tool sounds. Tranel (Tranel et al. 2003; Tranel et al. 2005) demonstrated modulation of the activation in the ventral inferior temporal cortex when naming sounds. Other neuroimaging studies showed a left posterior MTG focus for tool sounds and bilateral middle STG focus for animal sounds (Lewis et al. 2005; Doehrmann and Naumer 2008). In the latter 2 studies, the authors suggest that the region of middle STG is more sensitive to physical attributes of the sounds, whereas posterior MTG operates on higher processing levels and supports lexical-semantic processing.

For the 2 verbal modalities—spoken and written words, the searchlight technique revealed large clusters of category-discriminative voxels in bilateral middle and posterior STG—a region that supports lexical-phonological processing (Hickok and Poeppel 2007). Although the discriminative patterns that resulted from the group-level analysis seem to be right

dominant, this is likely due to the fact that the responses were more variable across participants in the left rather than in the right hemisphere.

The novelty of our research lies in demonstrating that the semantic category of a presented stimulus can be decoded not only using different input modalities, but also by generalizing the decoding across modalities. Our cross-modal classification results directly demonstrate that the left ITG, FG, and pMTG display a common pattern of activation for concrete objects, regardless of whether objects are presented visually, auditory, or verbally. This implies that neuronal circuits in these regions support the representation of semantic knowledge by integrating information originating from different input streams.

Shinkareva et al. (2011) applied a multivariate classification approach to identify the category of presented stimuli (dwellings vs. tools) and demonstrated the ability to generalize the classifier between 2 modalities (pictures and written words). To our knowledge, it is the only previously published study that showed generalization of a classifier across modalities. A direct comparison between the study of Shinkareva et al. (2011) and the present study is problematic as different signal analysis approaches were used. In Shinkareva et al. (2011), the modalities were compared at a level of distributed patterns of activations, whereas the present study examined local patterns. By analyzing a single brain region at a time, Shinkareva et al. (2011) revealed a large number of regions in which voxels were diagnostic for the semantic category in a similar way across the 2 modalities. On contrary to these results, in the present study, the whole-brain searchlight analysis identified a very distinctive cluster of common activation for the different modalities in the left ITG (Fig. 3).

The integrative role of the left ITG and ventrolateral occipitotemporal cortex has been suggested by previous studies. The left fusiform region supports complex visual processing and, for a long time, it has been assumed that the visual differences between stimuli of different categories fully account for the category-related patterns of activation in this region. However, accumulation of evidence in recent years suggests that this region is sensitive to semantic manipulations (Vandenberghe et al. 1996; Moore and Price 1999; Martin and Chao 2001; Price et al. 2003; Devlin et al. 2005; Wheatley et al. 2005; Martin 2007; Birn et al. 2010; Van Doren et al. 2010).

A number of authors considered left posterior MTG as a key node in the system of storage and retrieval of lexical semantic information in the brain (Hagoort et al. 2009; Snijders et al. 2009; Binder and Desai 2011). In a recent study, Wei et al. (2012) demonstrated that the amplitude of spontaneous low-frequency fluctuations in the BOLD signal in the left pMTG is highly correlated with individual variability in conceptual processing efficiency, assessed with behavioral tests. The behavioral tests included picture and sound object naming and picture associative matching. The left pMTG was the only brain region in which the spontaneous neuronal activity was significantly correlated with performance in all tests, regardless of stimulus modality (Wei et al. 2012).

Studies on audio–visual integration further support the suggested integrative role of the left posterior MTG and FG. This research focuses on how information from several sensory modalities is integrated to form a coherent whole. In a natural perceptual situation, a complex visual stimulus is paired with a matching auditory counterpart, for instance, the picture of a car and its corresponding engine sound. Neuroimaging studies have intended to capture this process of audio–visual integration, either by contrasting unimodal and bimodal stimuli or by contrasting congruent and incongruent bimodal stimuli. In a number of studies, posterior MTG has been shown to be important for integrating multimodal information about complex stimuli (Beauchamp et al. 2004; Naumer et al. 2009). Activation in the left FG has also been found in studies on audio–visual integration (Adams and Janata 2002), as reviewed by Doehrmann et al. (2008).

The generalization tests also revealed a number of frontal areas involved in the semantic task, regardless of input modality, which indicates abstracting away from low-level perceptual features to higher-level conceptual processing. These results are consistent with previous research. Lesions of the dorsal and medial frontal lobe cause transcortical motor aphasia (Luria and Tsvetkova 1968; Alexander 1997), a syndrome with characteristic nonfluency in spontaneous speech and word-finding problems, especially in cases when a large set of responses is possible (Robinson et al. 1998). Patients are often able to name objects relatively normally but are unable to generate a list of words within a category when no cue is presented (Luria and Tsvetkova 1968; Freedman et al. 1984). Neuroimaging studies on semantic fluency, which is an ability to generate word lists, have shown activation in the superior–anterior part of the MFG (Birn et al. 2010).

The present results strongly agree with a hierarchical neuroanatomical model of semantic processing as recently proposed by Binder and Desai (2011). In short, this model describes semantic processing as an interactive continuum, where low-level modality-specific representations interact with higher-level convergence zones. The convergence zones in the inferior parietal lobe and much of the ventral and lateral temporal lobe bind representations from multiple modalities and encode an abstract, schematic level of concept representation. Binder further proposes that prefrontal regions similar to those identified in the current study control top–down activations and are necessary for self-guided, goal-directed semantic retrieval (Binder et al. 2009; Binder and Desai 2011).

Successful decoding of semantic category independent of input modality also points toward the possibility of developing a BCI which is driven exclusively by concept selection. The experimental scenario of the final analysis, where we assessed whether the classifier could identify semantic category in free recall, is close to a desired real-life situation of BCI usage. In this analysis, we compared decoding performances of different classifiers: each classifier was trained on the data from 1 modality and tested on the free recall data, and voxels that allowed for generalization across 4 modalities were used. The above-chance performances obtained in this test indicate the possibility of decoding semantic information from the brain, during nondirected free recall, in the absence of any acoustic or visual cues. The improvement of prediction accuracy when trained on the combination of all modalities confirms that the patterns of the neuronal activations captured by the generalization test reflect conceptual representations that abstract away from specific sensory attributes of presented stimuli.

Summarizing, our study has shown that decoding of semantic category independent of input modality is a feasible

goal. Our results lead us to conclude that left inferior temporal cortex in combination with right middle and superior frontal cortex and left inferior frontal cortex could be a viable substrate for such decoding. The present results offer an opportunity to decode semantic information from brain signals and pave the way for understanding the mechanisms of storage and retrieval of semantic information. In this view, the present study is especially interesting given that categorical discrimination is one of the basic principles of organization of semantic knowledge in humans (Mandler 2004).

Supplementary Material

Supplementary material can be found at: http://www.cercor. oxfordjournals.org/.

Notes

References

Adams RB, Janata P. 2002. A comparison of neural circuits underlying auditory and visual object categorization. Neuroimage. 16:361–377.

Alexander MP. 1997. Aphasia: clinical and anatomic aspects. In: Feinberg TE, Farah ME, editors. Behavioral neurology and neuropsychology. New York: McGraw-Hill. p. 133–149.

Beauchamp MS, Lee KE, Argall BD, Martin A. 2004. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron. 41:809–823.

Binder JR, Desai RH. 2011. The neurobiology of semantic memory. Trends Cogn Sci. 15:527–536.

Binder JR, Desai RH, Graves WW, Conant LL. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex. 19:2767–2796.

Birn RM, Kenworthy L, Case L, Caravella R, Jones TB, Bandettini PA, Martin A. 2010. Neural systems supporting lexical search guided by letter and semantic category cues: a self-paced overt response fMRI study of verbal fluency. Neuroimage. 49:1099–1107.

Bright P, Moss H, Tyler LK. 2004. Unitary vs multiple semantics: PET studies of word and picture processing. Brain Lang. 89:417–432.

Brumberg JS, Nieto-Castanon A, Kennedy PR, Guenther FH. 2010. Brain-computer interfaces for speech communication. Speech Commun. 52:367–379.

Buckner RL, Koutstaal W, Schacter DL, Rosen BR. 2000. Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. Brain. 123:620–640.

Chao LL, Haxby JV, Martin A. 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat Neurosci. 2:913–919.

Chao LL, Weisberg J, Martin A. 2002. Experience-dependent modulation of category-related cortical activity. Cereb Cortex. 12:545–551.

Cox DD, Savoy RL. 2003. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage. 19:261–270.

Dasalla CS, Kambara H, Sato M, Koike Y. 2009. Single-trial classification of vowel speech imagery using common spatial patterns. Neural Netw. 22:1334–1339.

Devlin JT, Rushworth MF, Matthews PM. 2005. Category-related activation for written words in the posterior fusiform is task specific. Neuropsychologia. 43:69–74.

Doehrmann O, Naumer MJ. 2008. Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. Brain Res. 1242:136–150.

Doehrmann O, Naumer MJ, Volz S, Kaiser J, Altmann CF. 2008. Probing category selectivity for environmental sounds in the human auditory brain. Neuropsychologia. 46:2776–2786.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K. 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage. 25:1325–1335.

Formisano E, De Martino F, Valente G. 2008. Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. Magn Reson Imaging. 26:921–934.

Freedman M, Alexander MP, Naeser MA. 1984. Anatomic basis of transcortical motor aphasia. Neurology. 34:409–417.

Hagoort P, Baggio G, Willems RM. 2009. Semantic unification. In: Gazzaniga MS, editor. The Cognitive Neurosciences. Boston (MA): MIT Press. p. 819–836.

Hauk O, Davis MH, Kherif F, Pulvermüller F. 2008. Imagery or meaning? Evidence for a semantic origin of category-specific brain activity in metabolic imaging. Eur J Neurosci. 27:1856–1866.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 293:2425–2430.

Haynes JD, Rees G. 2006. Decoding mental states from brain activity in humans. Nat Rev Neurosci. 7:523–534.

Hickok G, Poeppel D. 2007. The cortical organization of speech processing. Nat Rev Neurosci. 8:393–402.

Just MA, Cherkassky VL, Aryal S, Mitchell TM. 2010. A neurosemantic theory of concrete noun representation based on the underlying brain codes. PLoS One. 5:e8622.

Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. Nat Neurosci. 8:679–685.

Kircher T, Sass K, Sachs O, Krach S. 2009. Priming words with pictures: neural correlates of semantic associations in a cross-modal priming task using fMRI. Hum Brain Mapp. 30:4116–4128.

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. Proc Natl Acad Sci USA. 103:3863–3868.

Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA. 2005. Distinct cortical pathways for processing tool versus animal sounds. J Neurosci. 25:5148–5158.

Luria AR, Tsvetkova LS. 1968. The mechanism of "dymanic aphasia". Foundations of Language. 4:296–307.

Mahon BZ, Caramazza A. 2009. Concepts and categories: a cognitive neuropsychological perspective. Annu Rev Psychol. 60:27–51.

Mandler JM. 2004. Thought before language. Trends Cogn Sci. 8:508–513.

Marinkovic K, Dhond RP, Dale AM, Glessner M, Carr V, Halgren E. 2003. Spatiotemporal dynamics of modality-specific and supramodal word processing. Neuron. 38:487–497.

Martin A. 2007. The representation of object concepts in the brain. Annu Rev Psychol. 58:25–45.

Martin A, Chao LL. 2001. Semantic memory and the brain: structure and processes. Curr Opin Neurobiol. 11:194–201.

Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA. 2008. Predicting human brain activity associated with the meanings of nouns. Science. 320(5880):1191–1195.

Moore CJ, Price CJ. 1999. A functional neuroimaging study of the variables that generate category-specific object processing differences. Brain. 122:943–962.

Naumer MJ, Doehrmann O, Müller NG, Muckli L, Kaiser J, Hein G. 2009. Cortical plasticity of audio-visual object representations. Cereb Cortex. 19:1641–1653.

Noppeney U, Price CJ, Penny WD, Friston KJ. 2006. Two distinct neural mechanisms for category-selective responses. Cereb Cortex. 16:437–445.

Oostenveld R, Fries P, Maris E, Schoffelen JM. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci. 2011:156869.

Op de Beeck HP, Haushofer J, Kanwisher NG. 2008. Interpreting fMRI data: maps, modules and dimensions. Nat Rev Neurosci. 9:123–135.

Patterson K, Nestor PJ, Rogers TT. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. Nat Rev Neurosci. 8:976–987.

Pereira F, Mitchell T, Botvinick M. 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage. 45:S199–S209.

Pobric G, Jefferies E, Lambon Ralph MA. 2010. Category-Specific versus Category-General Semantic Impairment Induced by Transcranial Magnetic Stimulation. Curr Biology. 20:964–968.

Poser BA, Versluis MJ, Hoogduin JM, Norris DG. 2006. BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. Magn Reson Med. 55:1227–1235.

Price CJ. 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. Ann N Y Acad Sci. 1191:62–88.

Price CJ. 2004. An Overview of Speech Comprehension and Production. In: Frackowiak RSJ, Friston KJ, editors. Human brain function. Amsterdam, Boston: Elsevier Academic Press. p. 517–533.

Price CJ, Winterburn D, Giraud AL, Moore CJ, Noppeney U. 2003. Cortical localisation of the visual and auditory word form areas: a reconsideration of the evidence. Brain Lang. 86:272–286.

Reddy L, Kanwisher N. 2007. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. Curr Biol. 17:2067–2072.

Riddoch MJ, Humphreys GW, Coltheart M, Funnell E. 1988. Semantic systems or system? Neuropsychological evidence re-examined. Cogn Neuropsychol. 5:3–25.

Robinson G, Blair J, Cipolotti L. 1998. Dynamic aphasia: an inability to select between competing verbal responses? Brain. 121(Pt 1):77–89.

Schneider TR, Engel AK, Debener S. 2008. Multisensory identification of natural objects in a two-way crossmodal priming paradigm. Exp Psychol. 55:121–132.

Shinkareva SV, Malave VL, Mason RA, Mitchell TM, Just MA. 2011. Commonality of neural representations of words and pictures. Neuroimage. 54:2418–2425.

Shinkareva SV, Mason RA, Malave VL, Wang W, Mitchell TM, Just MA. 2008. Using FMRI brain activation to identify cognitive states associated with perception of tools and dwellings. PLoS One. 3: e1394.

Snijders TM, Vosse T, Kempen G, Van Berkum JJ, Petersson KM, Hagoort P. 2009. Retrieval and unification of syntactic structure in sentence comprehension: an FMRI study using word-category ambiguity. Cereb Cortex. 19:1493–1503.

Tranel D, Damasio H, Eichhorn GR, Grabowski T, Ponto LL, Hichwa RD. 2003. Neural correlates of naming animals from their characteristic sounds. Neuropsychologia. 41:847–854.

Tranel D, Grabowski TJ, Lyon J, Damasio H. 2005. Naming the same entities from visual or from auditory stimulation engages similar regions of left inferotemporal cortices. J Cogn Neurosci. 17:1293–1305.

Van Doren L, Dupont P, De Grauwe S, Peeters R, Vandenberghe R. 2010. The amodal system for conscious word and picture identification in the absence of a semantic task. Neuroimage. 49:3295–3307.

Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RS. 1996. Functional anatomy of a common semantic system for words and pictures. Nature. 383:254–256.

Vandenbulcke M, Peeters R, Dupont P, Van Hecke P, Vandenberghe R. 2007. Word reading and posterior temporal dysfunction in amnestic mild cognitive impairment. Cereb Cortex. 17:542–551.

van Gerven M, Farquhar J, Schaefer R, Vlek R, Geuze J, Nijholt A, Ramsey N, Haselager P, Vuurpijl L, Gielen S et al. 2009. The brain-computer interface cycle. J Neural Eng. 6:041001.

Vapnik VN. 2000. The nature of statistical learning theory. Statistics for Engineering and Information Science. New York: Springer.

Vindiola M, Wolmetz M. 2011. Mental encoding and neural decoding of abstract cognitive categories: a commentary and simulation. Neuroimage. 54:2822–2827.

Wei T, Liang X, He Y, Zang Y, Han Z, Caramazza A, Bi Y. 2012. Predicting conceptual processing capacity from spontaneous neuronal activity of the left middle temporal gyrus. J Neurosci. 32:481–489.

Wheatley T, Weisberg J, Beauchamp MS, Martin A. 2005. Automatic priming of semantically related words reduces activity in the fusiform gyrus. J Cogn Neurosci. 17:1871–1885.

Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. 2002. Brain–computer interfaces for communication and control. Clin Neurophysiol. 113:767–91.