

## Identifying bilingual semantic neural representations across languages

Augusto Buchweitz<sup>a,d,\*</sup>, Svetlana V. Shinkareva<sup>b</sup>, Robert A. Mason<sup>a</sup>, Tom M. Mitchell<sup>c</sup>, Marcel Adam Just<sup>a</sup>

<sup>a</sup> Center for Cognitive Brain Imaging, Carnegie Mellon University, Pittsburgh, PA, United States

<sup>b</sup> Department of Psychology, University of South Carolina, Columbia, SC, United States

<sup>c</sup> Machine Learning Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, United States

<sup>d</sup> Department of Modern Languages, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

### ARTICLE INFO

#### Article history:

Accepted 4 September 2011

Available online 5 October 2011

#### Keywords:

fMRI  
Bilingualism  
Pattern-based classification

### ABSTRACT

The goal of the study was to identify the neural representation of a noun's meaning in one language based on the neural representation of that same noun in another language. Machine learning methods were used to train classifiers to identify which individual noun bilingual participants were thinking about in one language based solely on their brain activation in the other language. The study shows reliable ( $p < .05$ ) pattern-based classification accuracies for the classification of brain activity for nouns across languages. It also shows that the stable voxels used to classify the brain activation were located in areas associated with encoding information about semantic dimensions of the words in the study. The identification of the semantic trace of individual nouns from the pattern of cortical activity demonstrates the existence of a multi-voxel pattern of activation across the cortex for a single noun common to both languages in bilinguals.

© 2011 Elsevier Inc. All rights reserved.

### 1. Introduction

The goal of this study was to explore whether it is possible to use brain activation to identify the semantic neural representation of individual nouns in bilinguals, regardless of the language used. Pattern-based classifiers were trained to identify which of 14 nouns a participant was considering based only on their observed fMRI brain activation in another language. To our knowledge, there are no neuroimaging studies that used machine-learning methods to identify patterns of brain activation associated with meaning representation across different languages. In this technique, a classifier was trained using fMRI data collected when words were presented in one language and was then tested to determine whether it could successfully decode the brain activation when given new fMRI images collected when words were presented in the other language. The logic of this experiment was that the classifier would be able to identify the brain activation across languages only if the words elicited similar semantic neural representations. This is quite different than showing a similar group-level activation pattern in the brain for semantic knowledge across languages, cross-language identification of neural patterns demonstrates that common semantic traces can be identified at the individual word level.

Bilingual researchers have long asked whether there is a unitary semantic representation across languages. It is understood that the

connections between words in L2 and the semantic representations of the first language (L1) strengthen as proficiency in L2 increases. Also, the age at which words are learned in L2 has an influence on the strength of semantic representations, (see [Hernandez and Li \(2007\)](#) for a review). The participants in the present study were late L2 learners of English immersed in the L2 environment.

The central characteristic of the pattern-based classification approach applied to fMRI is the identification of a multivariate (multi-voxel) pattern of activation levels that collectively identifies the neural response to a stimulus. This pattern of activation represents an identifiable structure of multiple brain areas that are consistently activated in the tasks. Studies with pattern-based classification and brain imaging have shown that it is possible to identify what object a person is thinking about by training classifiers on brain activation ([Cox & Savoy, 2003](#); [Hanson, Matsuka, & Haxby, 2004](#); [Haxby et al., 2001](#); [Haynes & Rees, 2006](#); [Mourao-Miranda, Bokde, Born, Hampel, & Stetter, 2005](#); [O'Toole et al., 2007](#); [Shinkareva et al., 2008](#)). Recent studies of pattern-based classification of fMRI data have also been able to decompose the semantic dimensions of the brain representation based on the multiple brain areas used for the identification of thoughts ([Just, Cherkassky, Aryal, & Mitchell, 2010](#); [Mitchell et al., 2008](#)). One of these semantic dimensions identified is associated with the graspability of physical objects and can be associated with verbs of perception and action ([Mitchell et al., 2008](#)).

Previous brain imaging studies have shown that similarity in bilingual brain activation for language processes increases generally as a function of language skill ([Illes et al., 1999](#); [Isel,](#)

\* Corresponding author at: Universidade Federal do Rio Grande do Sul, Instituto de Letras, Department of Modern Languages, Bento Gonçalves 9500, Porto Alegre 91540, RS, Brazil.

E-mail address: [abuchweitz@gmail.com](mailto:abuchweitz@gmail.com) (A. Buchweitz).

Baumgaertner, Thrän, Meisel, & Büchel, 2010). We hypothesized that the semantic neural representations of nouns would have a high level of cross-language commonality for the proficient bilinguals in this study. This neural representation would help reveal a semantically-rich mental content of the representation in either language. In other words, the pattern-based classification would reveal that the neural representation can be related to the semantic properties of the nouns (e.g. their purpose, how people interact with them). The study also aimed to uncover the location of the voxels that contained information for the identification of nouns. If cross-language classification of words is successful, the pattern of brain activation identified would suggest that similar semantic dimensions are elicited by words in L1 and L2, and that there would be a top-down (language-independent) activation of multiple areas associated with the semantic properties of the nouns.

## 2. Methods

### 2.1. Participants

Eleven right-handed speakers of Portuguese as a first language (three female), mean age 29.9 years ( $SD = 5.74$ ; range = 20–40 years), participated in the study. Each participant gave signed informed consent approved by the University of Pittsburgh and Carnegie Mellon University Institutional Review Boards.

Mean age at beginning of L2 contact and learning was 13.08 years ( $SD = 3.1$ ; range = 10–22 years) (late bilinguals (Paradis, 2003)). All participants had passed university-level English proficiency exams (e.g., TOEFL, IELTS) prior to beginning schooling in the US. To assess level of proficiency, participants filled out a language background questionnaire (adapted from Hasegawa, Carpenter, & Just, 2002). The questionnaire asked for self-ratings on reading and listening skills in English, on a scale of 1.0 (*poor*) to 5.0 (*excellent*). Bilinguals rated themselves as being highly proficient in reading in English ( $M = 4.25$ ; range = 3.0–5.0). Ten participants had been living in the United States for over 3 years, and one for less than a year. Ten of the participants were enrolled in graduate-level courses, and one in undergraduate, at Carnegie Mellon University, the University of Pittsburgh, or Duquesne University. All participants reported reading in English for at least 1 h a day. Though participants rated themselves as being highly proficient in reading in L2, it is likely that some if not most of the bilinguals in the study are unbalanced bilinguals (i.e. proficiency in L1 is superior to L2).

### 2.2. Experimental paradigm

Participants viewed words in Portuguese and English in two separate imaging acquisitions whose order was balanced across participants. The stimuli were concrete nouns from two semantic categories (tools and dwellings), with seven exemplars per category. The seven tools were hammer, screwdriver, saw, wrench, pliers, hatchet, and drill (in Portuguese: *martelo*, *chave de fenda*, *serra*, *chave de boca*, *alicate*, *machadinho*, and *furadeira*). The seven dwellings were palace, castle, shack, apartment, mansion, hut, and house (in Portuguese: *palácio*, *castelo*, *barraco*, *apartamento*, *mansão*, *cabana*, and *casa*). The 14 stimuli of each language were presented in six consecutive blocks of trials, each time in a different random permutation order (i.e., there were six presentations of each of the 14 words in each language).

The words were presented in white against a black background. Participants were instructed to silently read each word and consistently think of the same properties of the named concept during each presentation of the word. Each stimulus was presented for 3 s, followed by a 7 s rest period, during which the participants

were instructed to clear their minds and fixate on an X displayed in the center of the screen. There were six additional presentations of a fixation, 21 s each, distributed across the session, to provide a baseline measure of activation. A schematic representation of the paradigm is shown in Fig. 1.

### 2.3. fMRI procedure

Functional images were acquired on a Siemens Allegra 3.0T scanner at the Brain Imaging Research Center of Carnegie Mellon University and the University of Pittsburgh (gradient echo EPI pulse sequence; TR = 1000 ms, TE = 30 ms, and a 60° flip angle). Sixteen 5-mm thick oblique-axial slices were imaged (1-mm gap between slices). The acquisition matrix was  $64 \times 64$  with  $3.125 \times 3.125 \times 5$ -mm voxels.

### 2.4. fMRI data processing and analysis for machine learning

Initial data processing was performed with SPM2 (Wellcome Department of Imaging Neuroscience, UK). The data were corrected for slice timing, motion, and linear trend, and were normalized into MNI space without changing voxel size ( $3.125 \times 3.125 \times 6$  mm). Data were temporally smoothed with a high-pass filter using a 190-s cutoff. The data were normalized to the MNI template brain image using a 12-parameter affine transformation. Analyses of a single brain region at a time used region definitions derived from the Anatomical Automatic Labeling (AAL) system (Tzourio-Mazoyer et al., 2002).

The percent signal change (PSC) relative to the fixation condition was computed at each voxel for each fMRI image. The mean PSC (MPSC) of the four images acquired within a 4-s window (shifted 4 s from stimulus onset) provided the main input measure for the machine learning classifiers. This means that the classifier used the average of the four images for each of the six presentations for one word in one language. The PSC data for each noun presentation (a single 3D fMRI image for each of the six presentations) were further normalized to have mean zero and variance one.

### 2.5. Machine learning methods

Classifiers were trained to identify cognitive states associated with thinking about the properties of each noun in each language, using the evoked patterns of functional activity (mean PSC). Classifiers were functions  $f$  of the form:  $f: \text{mean\_PSC} \rightarrow Y_j$ ,  $j = \{1, \dots, m\}$ , where  $Y_j$  was one of the 14 nouns, and where  $\text{mean\_PSC}$  was a vector of mean PSC voxel activations. To reduce the dimensionality

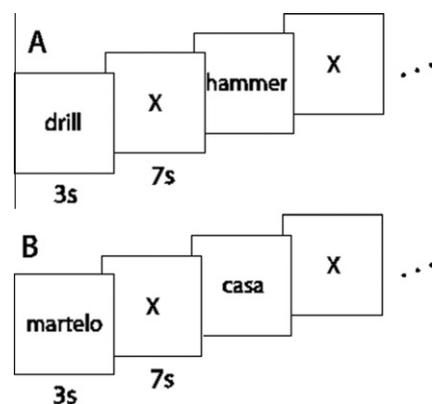


Fig. 1. Schematic representation of the experimental paradigm for the (A) English and (B) Portuguese acquisitions.

of the data, relevant features (voxels) were extracted from the training set prior to classification (see Section 2.6). A classifier was built from the training set (data in one language, e.g., Portuguese) using the selected features and was evaluated on the left-out test set (data in the other language, e.g., English) to ensure unbiased estimation of the classification error.

## 2.6. Feature selection

For the classification of brain activation *across languages*, the training set was entirely disjoint from the test set. The voxels were identified whose responses were the most stable over the set of training presentations of the nouns in one language and within each individual participant. There was no between-subject classification of data. The 120 most stable voxels were selected where voxel stability was computed as an average pairwise correlation between 14 noun vectors across the training presentations. For the training set in the classification of words *within language*, we used the MPSC images of four of the six presentations of each word in one language.

Maximizing accuracy is the appropriate goal in the scientific context of attempting to identify the content of a participant's thought. Our previous explorations of several feature selection methods and classifiers indicated only a small number of voxels need to be characterized to achieve good classification accuracy. The best classification rank-accuracies for fMRI data, as a function of the number of voxels used in feature selection, are obtained with approximately 120 stable voxels (Just et al., 2010; Mitchell et al., 2008; Shinkareva et al., 2008). Using more voxels (up to 2000) provided little accuracy gain in our previous studies.

## 2.7. Classification

We used the Gaussian Naïve Bayes (GNB) pooled variance classifier for classification purposes (Mitchell, 1997). It is a generative classifier that models the joint distribution of a class  $Y$  (e.g., tools or dwellings) and attributes (voxels), and assumes the attributes  $X_1, \dots, X_n$  are conditionally independent given  $Y$ . The classification rule is:

$$Y \leftarrow \underset{y_j}{\operatorname{argmax}} P(Y = y_j) \prod_i P(X_i | Y = y_j), j = 1, 2, \dots, m$$

In this experiment, all classes were equally frequent. Rank accuracy was used to evaluate 14-class classification (i.e. classification of the 14 verbs). To evaluate the significance of obtained rank accuracies, we performed random permutation tests (10,000 permutations) for each type of classification, and reported accuracies with  $p < .05$ . The rank accuracy of the classification performance was computed as the normalized rank of the correct label in the classifier's posterior-probability-ordered list of classes. For example, if the classification were operating at chance level, one would expect a mean normalized rank accuracy of 0.50, indicating that the correct word appeared on average between the 7th and 8th position in the classifier's output of a ranked list of 14 items. Use of rank-accuracy for describing a machine-learning classifier's accuracy with fMRI data is also described elsewhere (Just et al., 2010; Mitchell et al., 2008; Shinkareva et al., 2008).

Classification results for the within-language classifier were evaluated using k-fold cross-validation. The classifier was tested on the mean of the two left-out MPSC images of each word. This procedure was reiterated for all 15 possible combinations (folds) of leaving out two presentations. For the within-language classification, a rank accuracy was obtained for each fold, and these rank accuracies were averaged across folds, producing a single value characterizing the prediction accuracy for each word. The mean

accuracy across items (words) was then computed. The partitioning into four blocks of training trials and two blocks of test trials (the latter two being averaged) was determined by seeking maximum accuracy in previous studies. Averaging over two test trials simply reduces the noise and raises the accuracy by 3–5%.

## 2.8. Cross-validation

The classification across languages always left out the data for the to-be classified language and trained the classifier on the data for the other language (e.g., left out the data for nouns in English and trained the classifier on the data for nouns in Portuguese). The classifier was tested on the mean of the left-out six presentations of each word in the other, to-be classified language. The rationale for averaging over test trials is that it simply reduces the noise in the fMRI data and improves the classification accuracy.

To classify individual nouns within one language, two of the six MPSC images of each noun in one language were selected separately in each fold. Next, an average of these two presentations was computed and served as the test set. The remaining four MPSC images constituted the training set and were used for both the feature selection and the training of a classifier. The classifier was trained using separate noun labels. The trained classifier was used to predict the class of each test noun. This folding procedure resulted in 15 possible folds for within-language classification, and corresponding classification accuracies were averaged across folds.

## 2.9. Location of predictive voxels

The locations of predictive voxels are reported for the 120 most stable voxels used in the training set for cross-language classification. Anatomically-defined ROIs were used for counting the number of stable voxels in each brain lobe. The ROIs were defined using the parcellation proposed by Tzourio-Mazoyer et al. (2002).

## 2.10. fMRI analyses for SPM group contrasts

The data were analyzed using SPM2 (Wellcome Department of Cognitive Neurology). Images were corrected for slice acquisition timing, motion-corrected, normalized to the Montreal Neurological Institute (MNI) template, resampled to  $2 \times 2 \times 2$  mm voxels, and smoothed with an 8-mm Gaussian kernel to decrease spatial noise. Statistical analysis was performed on individual and group data by using the general linear model as implemented in SPM2 (Friston et al., 1995). The model for each participant included regressors for each of the two categories of interest (tools and dwellings), in each language, and for the fixation condition, convolved with the canonical SPM2 hemodynamic response function. The rest intervals between presentations of words were not explicitly modeled. To compare the distribution of activation across the tools and dwellings experimental conditions, group  $t$ -test analyses were performed using a random-effects model (Friston, Zarahn, Josephs, Henson, & Dale, 1999) using words versus fixation contrast images (one per participant, per contrast). All  $t$ -maps in each contrast were calculated across the entire cortical volume, thresholded at an uncorrected height threshold of  $p < .001$  and an extent threshold of 6 voxels. Statistical maps were superimposed on the high-resolution, normalized, T1-weighted, SPM2 individual template image for viewing. Labels for coordinates of activation were confirmed in MNI space (Tzourio-Mazoyer et al., 2002).

## 3. Results

The results show that it is possible to identify the semantic neural representation of a noun in one language based on the brain

activation for that same noun in another language in late bilinguals. The accuracy of the pattern-based classification also demonstrates that the classifier was comparably able to identify the neural representation based on brain activation in L1 or L2. The locations of the most stable voxels for the identification of brain activation across languages included cortical areas associated with thinking about the semantic dimensions associated with the physical objects (i.e., manipulation of tools and shelter). The location of the stable voxels suggests that the identification of brain activity across language was based on a semantically-rich, category-specific representation of the semantic properties of the nouns.

### 3.1. Classification of brain activity for words across languages

For the classification of brain activation from English to Portuguese (Eng–Pt), a classifier for each participant was trained on the brain activation for English words (L2) to identify the brain activation for the Portuguese words (L1). The mean rank accuracy for Eng–Pt single word classification was .68 ( $SD = 0.11$ ). The highest Eng–Pt classification rank accuracy obtained for a single participant was .82 (participant # 4). Our classifiers were also able to identify brain activation in the other direction, Portuguese to English (Pt–Eng). A classifier for each participant was trained on the brain activation for Portuguese words to identify the brain activation for English words. The mean rank accuracy for Pt–Eng single word classification was .72 ( $SD = 0.08$ ). The highest classification rank accuracy obtained for a single participant was .89 (participant # 9) (Fig. 2).

### 3.2. Locations of the voxels that identify brain activation across languages

The stable voxels were located in multiple areas of the brain that indicate some degree of cross-language commonality of the neural representation of the semantic dimensions of nouns. The locations of the stable voxels indicate that the fMRI data used for cross-language identification of brain activity may be encoding category-specific semantic properties of tools and dwellings. The overall characteristics of stable voxel locations were: (1) left post-central and supramarginal gyri, inferior and superior parietal lobes, and cortical areas that are part of the language network of the brain (left inferior frontal gyrus and left posterior superior

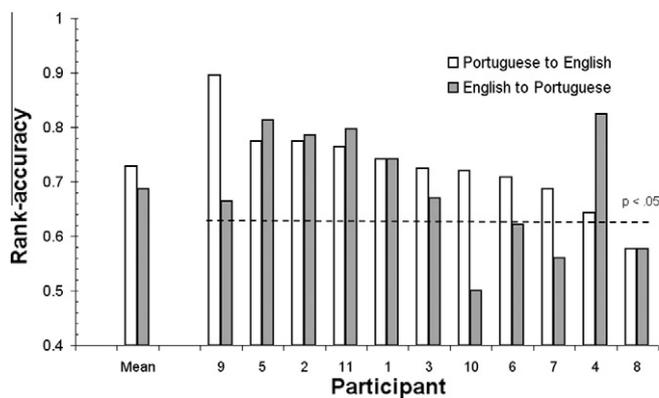
temporal lobe); (2) primarily located in the left hemisphere of the brain; and (3) primarily located in the frontal and occipital lobes (Fig. 3; Table 1).

The locations of some of the stable voxels used for the cross-language classification have been previously associated with the semantic neural representation of manipulable, physical objects and shelter (dwellings) (Just et al., 2010). English and Portuguese stable voxels overlapped with areas of the brain reported in association with these two semantic dimensions of concrete nouns. The semantic dimension of *manipulation* of objects was associated with brain activation for thinking about using objects, such as *pliers* and *keys*, in left supramarginal and postcentral gyri clusters (Just et al., 2010). These clusters overlapped with supramarginal and postcentral gyri voxels found in the stable voxels for the present study. Another semantic dimension reported by Just et al. (2010) was *shelter*, which was associated with the neural semantic representation of nouns such as *apartment* and *house*. This dimension of the semantic representation was associated with voxels in bilateral parahippocampal gyrus. The left parahippocampal gyrus location also matched the location of stable voxels found in the present study (Fig. 4).

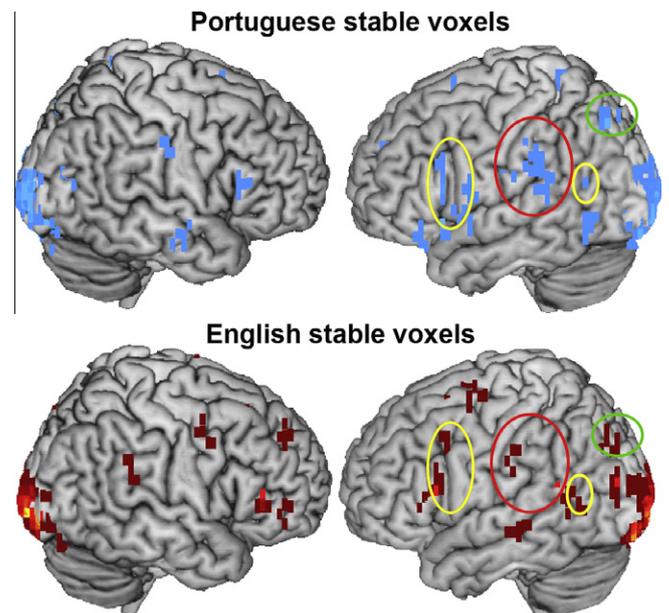
### 3.3. Classification of brain activity for words within languages

Our classifiers were also able to identify the brain activation for nouns within languages. For within-Portuguese classification of words (within participants), the mean rank accuracy was .63 ( $SD = 0.06$ ). For within-English classification of words (within participants), the mean rank accuracy was .60 ( $SD = 0.08$ ). The classification within language confirms the ability to identify thoughts associated with physical objects in first language of participants (Mitchell et al., 2008; Shinkareva et al., 2008). In addition to that ability, we were able to classify brain activation in the second language of the participants.

Participants 1 through 9 and participant 11 had significant rank accuracies (the non-significant rank accuracy for participant 10



**Fig. 2.** Cross-language identification of neural representations for nouns: identification accuracies for the classifier trained on the brain activation for words for one participant used to predict the same words in the other language in the same participant. Classification for Eng–Pt is  $M = .68$ ;  $SD = 0.11$ ; and for Pt–Eng is  $M = .72$ ;  $SD = 0.08$ . Reliable accuracies (rank-accuracy  $> .63$ ;  $p < .05$ ) were obtained for seven participants in English-to-Portuguese (Eng–Pt) cross-language classification, and for 10 participants in Portuguese-to-English (Pt–Eng) classification. Results are rank-ordered from highest to lowest classification rank accuracies for the Pt–Eng classification (white bars).

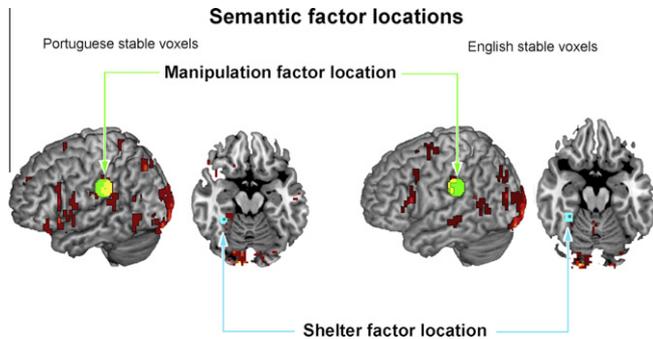


**Fig. 3.** Union of stable voxels for English words and for Portuguese words for all participants. Stable voxel clusters (minimum five stable voxels) show similarities in brain areas where the voxels used for classification of brain activation across languages were located. Yellow ellipses highlight areas in the language network (LIFG and left posterior superior temporal lobe). Red ellipses highlight areas in the postcentral gyrus. Green ellipses highlight areas in the left inferior parietal sulcus.

**Table 1**  
Distribution of stable voxels according to brain lobes.

	Portuguese	SE	English	SE
Frontal lobe	38.9	3.8	32.8	4.2
Temporal lobe	11.5	1.5	12.5	1.8
Parietal lobe	14.1	1.8	17.2	1.2
Occipital lobe	38.9	3.8	32.8	4.2

Stable voxel counts based on anatomical regions of interest. These anatomical ROIs were defined in the Montreal Neurological Institute (MNI) space using the parcellation proposed by Tzourio-Mazoyer et al. (2002) for the single participant MNI brain.

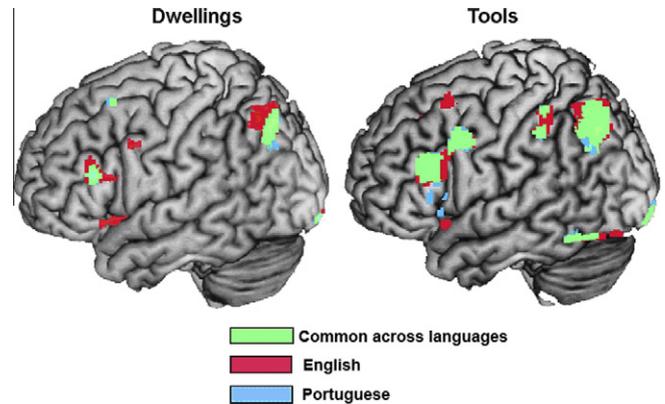


**Fig. 4.** Voxels used in pattern-based classification that encode the semantic representation of object manipulation and shelter in both Portuguese and English shown in red or yellow. Figure shows voxel clusters from the union of stable voxels across participants. Green circles highlight one of the manipulation factor areas in left supramarginal/postcentral gyrus reported by Just et al. (2010) (centroids are  $x = -60$ ,  $y = -30$ ,  $z = 34$ ; 51 voxels and a radius of 10.0 mm). Blue circles highlight one of the shelter factor areas in left fusiform/parahippocampal gyri reported by Just et al. (2010) ( $x = -32$ ,  $y = -42$ ,  $z = -18$ ; 26 voxels and a radius of 6.0 mm). Yellow areas indicate stable voxels in the current study.

was .51). The highest classification rank accuracy for Portuguese words was .72 (participant # 2). For within-English classification of words (within participants), the mean rank accuracy was .60 ( $SD = 0.08$ ) and the rank accuracies of eight of the 11 participants were reliably above chance. Participants 1 through 6 and participants 9 and 11 had significant rank accuracies (the non-significant rank accuracies were, for participant 7, .49, for participant 8, .47, and for participant 10, .50). The highest classification rank accuracy for English was .71 (participant # 3). The within-language results for the classification of words confirmed the ability to predict brain activity based on the activation associated with individual words. Within-language classification of brain activity also showed that the participants with the poorest classification results across languages had the poorest classification results for words in English (the second language).

### 3.4. Activation for tools and dwellings

In addition to reporting the locations of stable voxels for the pattern-based classification of brain activity within participants, we report the group-level activation for the tools and dwellings in English and Portuguese. The group-level brain activation shows remarkable similarities in the areas activated for tools and dwellings across languages. The brain activation for tools and dwellings in either English or Portuguese was mostly left-lateralized, and included the areas of the precentral and postcentral gyri, the language network, and inferior and posterior parietal lobe (Fig. 5; Tables 2 and 3). Fig. 5 shows the large areas of activation, especially for tools, that were common to both languages (shown in green). The common areas were smaller for dwellings, perhaps



**Fig. 5.** Common areas of brain activation for thinking about the properties of tools and dwellings in English and Portuguese. Group-level contrast for Dwellings > Fixation (left) and for Tools > Fixation (right) for English (red), for Portuguese (blue), and for the commonalities between the two languages (green);  $p < .001$  uncorrected;  $T = 4.14$ ; extent threshold = 6 voxels.

because there is less commonality in how some of the dwelling words (e.g. *mansion*) are represented in English and Portuguese. Note that the red and blue legends correspond to areas where only one of the languages showed reliable activation, but there were no reliable activation differences between languages. Thus there is very substantial neural commonality across languages but also some language-specific activation. Our methods and analyses intentionally focused on the commonalities, but the language differences, which are too small to interpret here, are open to further investigation.

## 4. Discussion

The study indicates that it is possible to identify the neural representation associated with semantic properties of words in one language based on the semantic brain representation of that same word in the other language. This result suggests that skilled, late bilinguals have language-independent patterns of neural representation for words; it is possible to identify which word they are thinking about using the semantic neural representation elicited by words in either language. The study also showed that the stable voxels used for classification purposes were located in areas that encode semantic-related information about tools and dwellings.

Though the participants in the study learned English as a second language at a later stage in their lives, it is possible to identify the word they were considering based solely on brain activation patterns in another language. Studies have shown that late bilinguals have more differences than early bilinguals in their brain activation for different language tasks (Kim, Relkin, Lee, & Hirsch, 1997; Perani & Abutalebi, 2005). Studies also show that the differences in brain activation associated with semantic representations can be modulated by second language skill and proficiency rather than age of acquisition (Illes et al., 1999; Isel et al., 2010). Moreover, bilingual brain activation is also modulated by the type of language task (e.g. tasks that tap into syntax, language comprehension, or semantics (e.g. Wartenburger et al., 2003)). The issue of how age of acquisition and skill modulate brain activation remains a hot topic in brain imaging studies of bilingual participants. Behavioral studies, in turn, show that bilinguals establish a conceptual mapping of the meaning of words that does not rely on the translation of words (*lexical mapping*) to the first language (Kroll & de Groot, 1997). For example, proficient bilinguals are equally able to name pictures and categorize word exemplars (e.g., a *hammer* is a *tool*) in L1 and L2 (Caramazza & Brones, 1980; Potter, So, Von Eckhardt, & Feldman, 1984), and they suffer a Stroop effect

**Table 2**  
Activation for tools in English and tools in Portuguese compared with fixation.

Tools in English	Voxels	T (11)	MNI		
			x	y	z
<i>Frontal</i>					
L inferior frontal gyrus	481	7.49	-54	26	24
L precentral gyrus					
L middle frontal gyrus	41	6.73	-26	16	56
L superior frontal gyrus	11	6.69	-22	14	-14
R middle frontal gyrus	10	6.61	42	34	28
R inferior frontal gyrus					
L supp. motor area	154	5.72	-2	20	48
L medial frontal gyrus					
R middle frontal gyrus	13	5.24	32	50	32
L inferior frontal gyrus	29	4.80	-46	16	-10
<i>Parietal</i>					
L inferior parietal lobe	758	10.34	-34	-70	42
L superior parietal lobe					
L middle occipital lobe					
L angular gyrus					
L inferior parietal lobe	95	5.78	-46	-38	44
L postcentral gyrus					
R angular gyrus	56	5.22	42	-64	42
<i>Occipital</i>					
L lingual gyrus	355	8.78	-14	-40	0
L calcarine					
L parahipp. gyrus					
L fusiform gyrus	97	8.45	-42	-54	-16
L inferior temporal gyrus					
L lingual gyrus	58	6.49	-16	-100	-14
L calcarine					
L fusiform gyrus	8	4.47	-38	-80	-18
<i>Subcortical</i>					
R caudate	63	7.08	16	22	-4
<b>Tools in Portuguese</b>					
Tools in Portuguese	Voxels	T (11)	MNI		
<i>Frontal</i>					
L inferior frontal gyrus	290	7.23	-54	26	24
L precentral gyrus	202	6.38	-50	4	38
L superior frontal gyrus (orb)	8	5.57	-24	14	-14
L insula					
L supp. motor area	95	5.55	-4	22	46
L superior frontal gyrus					
R superior frontal gyrus	10	5.16	14	22	42
R medial frontal gyrus					
R supp. motor area					
R middle frontal gyrus	6	5.07	30	52	28
L inferior frontal gyrus	202	6.38	-50	24	8
L inferior frontal gyrus	6	4.38	-48	20	-2
<i>Parietal</i>					
L inferior parietal lobe	553	8.53	-34	-70	42
L middle occipital lobe					
L superior parietal lobe					
L angular gyrus					
L inferior parietal lobe	194	5.45	-24	-46	40
L postcentral gyrus	15	4.42	-48	-38	58
<i>Occipital</i>					
L fusiform gyrus	99	9.21	-42	-54	-16
L inferior temporal gyrus					
L lingual gyrus	451	8.43	-10	-70	8
L hippocampus					
L lingual gyrus	88	7.05	-16	-100	-14
L calcarine					
<i>Subcortical</i>					
R caudate	149	7.12	18	24	-4

Note: Clusters of voxels significant at  $p < .001$ , uncorrected, extent threshold = 6 voxels. Contrasts of all tools, separately in each language, and all dwellings, separately in each language, with fixation. Region labels apply to the entire extent of the cluster with peak maxima designated by first locale cited. *T*-values and MNI coordinates are for the peak activated voxel in each cluster.

from naming an ink color in one language when the word names a different color in another language (Chen & Ho, 1986; Mägiste, 1984). These studies indicate that there is a cross-language

**Table 3**  
Activation for dwellings in English and dwellings in Portuguese compared with fixation.

Dwellings in English	Voxels	T (11)	MNI		
			x	y	z
<i>Frontal</i>					
L inferior frontal gyrus	176	6.65	-48	28	20
R precentral gyrus	16	6.55	48	-18	42
R postcentral gyrus					
L supp. motor area	130	5.30	-4	22	44
L superior frontal gyrus					
L middle frontal gyrus	8	5.05	-26	18	56
L superior frontal gyrus					
R inferior frontal gyrus	9	4.71	38	14	12
R insula					
L inferior frontal gyrus (orb)	39	4.59	-42	10	-8
L insula					
L precentral gyrus	15	4.49	-50	6	40
<i>Parietal</i>					
L inferior parietal lobe	293	6.20	-30	-70	40
L superior parietal lobe					
<i>Temporal</i>					
L hippocampus	56	6.89	-18	-38	0
L lingual					
L parahipp. gyrus					
<i>Occipital</i>					
L lingual gyrus	29	7.20	-14	-100	-14
L calcarine					
L lingual gyrus	211	6.06	-8	-58	-2
<b>Dwellings in Portuguese</b>					
Dwellings in Portuguese	Voxels	T (11)	MNI		
<i>Frontal</i>					
L inferior frontal gyrus	75	5.51	-54	28	22
R precentral gyrus	13	5.46	50	-8	36
R postcentral gyrus					
L middle frontal gyrus	12	5.11	-24	18	56
L superior frontal gyrus					
L supp. motor area	41	4.72	-4	22	46
L superior frontal gyrus					
<i>Parietal</i>					
L inferior parietal lobe	145	5.87	-30	-74	46
L inferior parietal lobe	12	4.91	-30	-36	42
L postcentral gyrus					
R precuneus	7	4.70	20	-48	12
<i>Temporal</i>					
L hippocampus	41	6.08	-18	-38	2
<i>Occipital</i>					
L lingual gyrus	229	8.94	-6	-56	-4
L lingual gyrus	31	6.69	-14	-100	-14
L calcarine					

Note: Clusters of voxels significant at  $p < .001$ , uncorrected, extent threshold = 6 voxels. Contrasts of all tools, separately in each language, and all dwellings, separately in each language, with fixation. Region labels apply to the entire extent of the cluster with peak maxima designated by first locale cited. *T*-values and MNI coordinates are for the peak activated voxel in each cluster.

association between the semantic representations of words in skilled bilinguals.

Consistently thinking about semantic properties for L2 words may have posed an additional difficulty in the study. There was a slight advantage for classifying brain activation based on stable voxels identified in the first language of participants. Our results also indicate that language proficiency may have some bearing on the ability to use brain activation in L2 to predict activation in L1: there was a positive (though not significant) correlation between accuracy of prediction using L2 brain activation data (to predict L1 brain activation) and language proficiency ( $r = 0.39$ ). The more proficient the bilingual, the better the prediction of L1 brain activation using L2 brain activation. This suggests that the more proficient the bilingual, the more similarities in the neural representation of the words. Cross-language identification based on

brain activation for the first language (Portuguese) was reliable in 10 of the 11 participants. The single participant whose classification accuracy was not reliable also showed non-reliable classification accuracy in the other language direction, i.e., based on brain activation for words in English. This participant simply was not able to consistently think about the semantic properties of the words in either language. Cross-language identification based on brain activation for the L2 (English) was reliable in seven of the 11 participants. The classification was not reliable for participants 6, 7, 8, and 10, who also did not show reliable classification of brain activation for words in English-only (within-language). These participants may have had difficulty with the task of consistently thinking about the semantic properties of words in their L2. The difference in classification results may also be related to what has been described as different levels of reliance on conceptual and lexical mappings between first and second languages for relatively fluent, unbalanced bilinguals (Kroll & Stewart, 1994).

In the field of bilingualism it is generally understood that there is a time-locked constraint to one's ability to master a second language (see Hernandez and Li (2007) for a review), which is known as the 'critical period hypothesis' (the later an L2 is learned, the harder it is to achieve comparable skill in L2 processing and production). The task of generating and retrieving semantic properties elicited by words in a second language may be less automatic for some late bilinguals. It is possible that some of the participants relied on word associations between the languages to retrieve semantic properties for words in the second language (see also Kroll & Stewart, 1994). This would suggest a lexical, rather than conceptual, mapping of words to meaning in the second language. Though the participants in the study were proficient readers of English, for some, the later acquisition of the L2 may have resulted in a weaker or less automatic link between words and their semantic properties.

Thinking about the properties of words in the first language provided better training and test set data for pattern-based classification than thinking about the properties of words in the second language. The words in Portuguese, the L1, are likely to be more automatically associated with their semantic properties than the words in English, the less-salient L2. A more automatic retrieval of the semantic dimensions of words in L1 may allow participants to be more consistent each time they had to think about the semantic properties of each noun. It remains to be seen whether the same, or better, classification results can be obtained with early bilinguals (i.e. bilinguals who acquire their second language during childhood).

One of the fundamental goals addressed by cognitive neuroscience studies of bilingualism is to determine the factors that modulate brain activation in L2 and L1 (and differences between the factors). The participants in this study were highly-schooled, proficient bilinguals who began to learn the L2 later in life, in their early to mid-teens (late bilinguals; Paradis, 2003). Despite this participant sample being a reasonable starting point, studies have shown that factors such as age of acquisition, proficiency in L2, and various characteristics of the L1 and L2 (such as the transparency of print-to-sound mappings in orthography) have been found to influence the activation in L2 comprehension (e.g., Buchweitz, Mason, Hasegawa, & Just, 2009; Illes et al., 1999; Isel et al., 2010; Kim et al., 1997). Due to the diversity of these factors, it will probably be necessary to first investigate bilinguals' brain activation within a limited number of points in this large space, attempting first to infer general principles, and then subjecting the principles to specific tests in a number of situations.

In the recent past, a growing literature of brain imaging studies of bilingualism, shows that age of acquisition and proficiency determine differences in brain activation between L1 and L2 (e.g. Illes et al., 1999; Kim et al., 1997), but also that age of acquisition

modulates differences in brain activation between L1 and L2 only when proficiency is not taken into account (Illes et al., 1999; Perani et al., 1998). Proficient bilinguals, regardless of their age of acquisition, show remarkably similar brain activation for L1 and L2 comprehension and semantic processing tasks. The differences in brain activation between L1 and L2 for late, proficient bilinguals appear in tasks that involve the use of syntactic rather than semantic knowledge (Hernandez & Li, 2007).

The present study is consistent with the studies that show similarities in brain activation for late, proficient bilinguals. Despite the fact that the participants learned English in their late teens or in their early adult life, the bases for the neurosemantic representation of nouns shows a substantial similarity adequate for us to identify what noun the bilingual was thinking about in their other language.

The motor and shelter properties in the representations of the nouns provide evidence of embodied cognition, a theoretical position holding that meaning representations contain perceptual and motor components corresponding to one's interaction with objects in the physical environment (e.g., Barsalou, 1999; Glenberg, 1997). The semantic representation of manipulation for these nouns included clusters in left postcentral gyrus and inferior parietal lobe. The shelter representation for these nouns included clusters in left parahippocampal gyrus. These cortical areas have been associated with encoding the representation of manipulable objects and objects denoting shelter (Just et al., 2010). The stable voxels were also located in areas associated with visual processes, such as inferior temporal gyrus, occipital lobe, and the calcarine sulcus. The calcarine sulcus has been associated with the generation of visual-mental imagery that gives rise to experiences, based on retrieving stored information (Klein, Paradis, Poline, Kossly, & Le Bihan, 2000).

In addition to the voxels in more anterior, higher-order, associative areas of the brain, there were stable voxels located in more posterior areas of the brain associated with visual processing. Among these voxels, those in the parietal lobe may encode information related to evoking the personal experience of being in, or moving about, the dwellings. Activation in the right inferior parietal area of the brain has been found to support egocentric movement while moving about in virtual reality towns (Maguire et al., 1998). It has also been associated with tool usage and grasping motions (Culham, Cavina-Pratesi, & Singhal, 2006; Culham & Valyear, 2006). The visual cortex has been associated with a distinctive categorical representation of visual objects in a study that also used word stimuli, whereas activation in areas of higher-order association (such as temporal and frontal lobe areas) was less distinctive among objects (Ruschemeyer, Pfeiffer, & Bekkering, 2010).

The nouns in the study represent common objects encountered in everyday life situations. All participants reported being familiar with the nouns and being able to generate a semantic representation of the noun, for example, how they would interact with it. It seems plausible that the reliable cross-language identification of brain activity would also be achieved with other common nouns that are encountered in everyday life, and that are learned by those who make practical, everyday use of two languages (e.g., fruits and vegetables, transportation, and furniture, among others).

Although our findings indicate the substantial degree of cross-language neural commonality of representation for the 14 concrete words that were used, the commonality for other types of words remains to be investigated. The limited stimulus word set used here was chosen precisely because one might expect minimal cross-language differences in neural representation for words like *hammer*, because the interaction with a hammer is similar across languages and cultures. A second rationale for the choice of stimulus items is that previous research has found that visual presentations of objects in the categories of tools and dwellings tend to activate systematic and differentiable brain areas. These

considerations raise the question of how common the neural representation might be across languages for less common and less tangible concepts. Some concepts, like *anger*, are less tangible than a hammer, but might still be common across languages and cultures. We furthermore suggest, based on our studies still in progress, that the activation patterns for emotions are highly systematic across individuals, at least within a language. However, for more culture-bound concepts, such as *freedom* or *honor*, the activation patterns might be expected to differ substantially across languages as used in different cultures. In fact, the experimental methods used in our study provide a way to find and possibly interpret cultural differences in lexical representations. Perhaps this is another way of saying that translation non-equivalence for certain types of words may be open to new types of neurosemantic research that may have the capability of pinpointing the nature of the non-equivalence.

To our knowledge, this is the first study to predict the brain activity associated with thinking about the properties of words in bilingual participants. Combined with earlier studies that show similarities in neural representations across picture/word stimuli (Mitchell et al., 2008; Shinkareva et al., 2008), the study suggests that the brain areas that decode semantic content are mostly unaffected by stimulus type (English, Portuguese, drawing, words) and are therefore very likely to be locations that code purely semantic content rather than surface level percepts. The study also shows that semantic processing is organized over a common network of areas, allowing reliable decoding of brain representations across languages. It is possible to identify which word a person is thinking about based on their representation for the same word in a different language.

## Acknowledgments

This research was supported by NIMH Grant MH029617, NSF Grant IIS-0835797, and the W. M. Keck Foundation. We would like to thank Vladimir Cherkassky, Vincent Malave, and Sandesh Aryal for their data analytic contribution to this work.

## References

- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Buchweitz, A., Mason, R. A., Hasegawa, M., & Just, M. A. (2009). Japanese and English sentence reading comprehension and writing systems: An fMRI study of first and second language effects on brain activation. *Bilingualism: Language and Cognition*, 12, 141–151.
- Caramazza, A., & Brones, I. (1980). Semantic classification by bilinguals. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 34(1), 77–81.
- Chen, H.-C., & Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 397–401.
- Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) “brain reading”: Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, 19, 261–270.
- Culham, J. C., Cavina-Pratesi, C., & Singhal, A. (2006). The role of parietal cortex in visuomotor control: What have we learned from neuroimaging? *Neuropsychologia*, 44(13), 2668–2684.
- Culham, J. C., & Valyear, K. F. (2006). Human parietal cortex in action. *Current Opinion in Neurobiology*, 16, 205–212.
- Friston, K., Ashburner, J., Frith, C., Poline, J.-B., Heather, J., et al. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, 2, 165–189.
- Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N., & Dale, A. M. (1999). Stochastic designs in event-related fMRI. *Neuroimage*, 10, 607–619.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1–55.
- Hanson, S. J., Matsuka, T., & Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: Is there a “face” area? *Neuroimage*, 23, 156–166.
- Hasegawa, M., Carpenter, P. A., & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *Neuroimage*, 15, 647–660.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., et al. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430.
- Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Review Neuroscience*, 7(7), 523–534.
- Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin*, 133(4), 638–650.
- Illes, J., Francis, W. S., Desmond, J. E., Gabrieli, J. D. E., Glover, G. H., et al. (1999). Convergent cortical representation of semantic processing in bilinguals. *Brain and Language*, 70, 347–363.
- Isel, F., Baumgaertner, A., Thrän, J., Meisel, J. M., & Büchel, C. (2010). Neural circuitry of the bilingual mental lexicon: Effect of age of second language acquisition. *Brain and Cognition*, 72(2), 169–180.
- 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.
- Kim, K. H. S., Relkin, N. R., Lee, K., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, 388, 171–174.
- Klein, I., Paradis, A. L., Poline, J. B., Kossly, S. M., & Le Bihan, D. (2000). Transient activity in the human calcarine cortex during visual-mental imagery: An event-related fMRI study. *Journal of Cognitive Neuroscience*, 12, 15–23.
- Kroll, J. F., & De Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In J. F. Kroll & A. M. B. De Groot (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 169–200). New Jersey: Lawrence Erlbaum.
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174.
- Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10(2), 304–315.
- Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S. J., Frith, C. D., et al. (1998). Knowing where and getting there: A human navigation network. *Science*, 280, 921–924.
- Mitchell, T. M. (1997). In C. L. Liu (Ed.), *Machine learning*. Boston: McGraw-Hill.
- Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, 320, 1191–1195.
- Mourao-Miranda, J., Bokde, A. L., Born, C., Hampel, H., & Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. *Neuroimage*, 28, 980–995.
- O’Toole, A., Jiang, F., Abdi, H., Penard, N., Dunlop, J. P., et al. (2007). Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *Journal of Cognitive Neuroscience*, 18, 1–19.
- Paradis, M. (2003). Differential use of cerebral mechanisms in bilinguals. In M. T. Banich & M. Mack (Eds.), *Mind, brain, and language: Multidisciplinary perspectives* (pp. 351–370). London: Lawrence Erlbaum.
- Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, 15, 202–206.
- Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., et al. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain*, 121, 1841–1852.
- Potter, M. C., So, K.-F., Von Eckhardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23, 23–38.
- Ruschmeyer, S.-A., Pfeiffer, C., & Bekkering, H. (2010). Body schematics: On the role of the body schema in embodied lexical-semantic representations. *Neuropsychologia*, 48(3), 774–781.
- Shinkareva, S. V., Mason, R. A., Malave, V. L., Wang, W., Mitchell, T. M., et al. (2008). Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS One*, 3(1), e1394.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15, 273–289.
- Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., et al. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, 37, 159–170.