Cross-modal representation of spoken and written word meaning in left pars triangularis

Antonietta Gabriella Liuzzi[a], Rose Bruffaerts[a,b], Ronald Peeters[c], Katarzyna Adamczuk[a], Emmanuel Keuleers[d], Simon De Deyne[e], Gerrit Storms[e], Patrick Dupont[a], Rik Vandenberghe[a,b]

[a]Laboratory for Cognitive Neurology, Department of Neurosciences, KU Leuven, Belgium; [b]Neurology Department, University Hospitals Leuven, 3000 Leuven, Belgium; [c]Radiology Department, University Hospitals Leuven, 3000 Leuven, Belgium. [d]Department of Experimental Psychology, Ghent University, B-9000 Ghent, Belgium; [e]Laboratory of Experimental Psychology, Humanities and Social Sciences Group, KU Leuven, Belgium.

Correspondence to: Rik Vandenberghe. Neurology Department, University Hospitals Leuven, Herestraat 49 bus 7003, B-3000 Leuven, Belgium. Tel: +3216344280, Fax: +3216344285. E-mail: `rik.vandenberghe@uz.kuleuven.ac.be`.

Abstract

The correspondence in meaning extracted from written versus spoken input remains to be fully understood neurobiologically. Here, in a total of 38 subjects, the functional anatomy of cross-modal semantic similarity for concrete words was determined based on a dual criterion: First, a voxelwise univariate analysis had to show significant activation during a semantic task (property verification) performed with written and spoken concrete words compared to the perceptually matched control condition. Second, in an independent dataset, in these clusters, the similarity in fMRI response pattern to two distinct entities, one presented as a written and the other as a spoken word, had to correlate with the similarity in meaning between these entities. The left ventral occipitotemporal transition zone and ventromedial temporal cortex, retrosplenial cortex, pars orbitalis bilaterally, and the left pars triangularis were all activated in the univariate contrast. Only the left pars triangularis showed a cross-modal semantic similarity effect. There was no effect of phonological nor orthographic similarity in this region. The cross-modal semantic similarity effect was confirmed by a secondary analysis in the cytoarchitectonically defined BA45. A semantic similarity effect was also present in the ventral occipital regions but only within the visual modality, and in the anterior superior temporal cortex only within the auditory modality. This study provides direct evidence for the coding of word meaning in BA45 and positions its contribution to semantic processing at the confluence of input-modality specific pathways that code for meaning within the respective input modalities.

1 Introduction

According to one of the most influential contemporary language models, the neural network underlying speech processing can be divided into a dorsal and a ventral stream (Hickok and Poeppel, 2004; Ueno et al., 2011). Both streams project into the inferior frontal gyrus (IFG) which has a crucial role in the interaction of these two streams (Saur et al., 2008; Rijntjes et al., 2012; Hamzei et al., 2016). The most natural word input modality are spoken words (Chafe and Tannen, 1987) but in many cultures writing can convey meaning as efficiently. Given the similarity in the meaningful messages conveyed, the path connecting written word input with word meaning presumably converges with that for auditory input at a given stage. The cognitive and neurobiological architecture of this confluence between written and spoken word input has been of longstanding interest to evolutionary and developmental neuroscience (Chafe and Tannen, 1987), neuropsychology (Allport and Funnell, 1981) and, more recently, functional imaging of the intact brain (Chee et al., 1999; Wagner et al., 2001; Booth et al., 2002; Homae, 2002; Constable et al., 2004; Gold et al., 2005a; Spitsyna et al., 2006). The study of the differences and commonalities in processing of meaning between different input-modalities (written or spoken words, pictures, ...) should not be confounded with the study of the effect of type of attributes of concrete entities (e.g. shape, sound, etc) (Vandenbulcke et al., 2006; Huth et al., 2016) nor with the study of the inner format of semantic representations (Caramazza et al., 1990; Barsalou, 2016).

fMRI activations during semantic processing that are common for written and spoken words (Chee et al., 1999; Wagner et al., 2001; Booth et al., 2002; Homae, 2002; Constable et al., 2004; Gold et al., 2005a; Spitsyna et al., 2006) (for review see Binder et al. (2009)) can arise for various reasons. Domain-general processes (e.g. common selection (Thompson-Schill et al., 1997) or control processes (Gold et al., 2005b)) may operate on written and spoken word meaning that is represented at a distance (Gold et al., 2005b; Hagoort, 2005). Or written words may be sounded out internally and the phonological operations associated with this process may give rise to apparent commonality with spoken word input. Thirdly, neuronal populations may code for the meaningful content of the words independently of the input modality in which the words were originally presented. A classical conjunction univariate analysis might reveal conjoint activation for these different reasons. Representational similarity analysis (RSA) (Kriegeskorte et al., 2008; Fairhall and Caramazza, 2013; Devereux et al., 2013) provides an opportunity to directly test the representational content and its dependence on input-modality. Depending on the behavioral matrix to which the fMRI matrix is compared and the extent of the stimulus set, RSA can reach a level of cognitive specificity and item-by-item granularity that cannot be attained by previous approaches to cross-modal processing (Homae, 2002; Kircher et al., 2009; Sass et al., 2009; Simanova et al., 2014). Among the cross-modal RSA

studies, the majority used written words together with pictures (Devereux et al., 2013). An advantage of using words only is that it avoids the perceptual confound induced by the covariance between visual characteristics of pictures and their meaning (Fernandino et al., 2015): Compared to pictures, the relation between word form and meaning is far more arbitrary.

In the current study, we determined RSA effects of semantic similarity for cross-modal pairs, written versus spoken concrete nouns. First, a set of regions was defined based on an univariate analysis using an explicit semantic task performed with written and with spoken words compared to a lower-level control condition with consonant letter strings and rotated spectrograms, respectively. In a subsequent independent experiment, we determined within this set of regions whether semantic similarity between pairs of words is reflected in the activity pattern despite differences in word input-modality, spoken versus written. We also ascertained that the similarity in activity patterns could not be explained by phonological or orthographic similarity between words. In order to obtain a more complete picture of how crossmodal semantic similarity effects relate anatomically to input-modality specific processing pathways, we additionally searched for semantic similarity effects within the written or the spoken word processing pathway.

2 Subjects and Methods

2.1 Participants

Eighteen subjects (12 women, 6 men), between 18 and 28 years, participated in a first fMRI experiment (univariate analysis). Twenty different subjects (14 women, 6 men), between 18 and 28 years, participated in a second, independent experiment, which was optimized for RSA (see below). All subjects were native Dutch speakers, right-handed, free of neurological or psychiatric history and had no hearing impairment. There was no overlap between the two subject groups. All the procedures were approved by the Ethics Committee of the University Hospital of Leuven.

A previous paper (Liuzzi et al., 2015) that exclusively focused on effects in the left perirhinal volume imported from Bruffaerts et al. (2013b) relied on the same dataset as the second experiment.

2.2 Stimuli

A particular spoken or a particular written word that is repeatedly presented in the fMRI experiment will be called a "stimulus" and the concept referred to by that word regardless of input modality will be called an "entity".

In both experiments, the stimulus set and the properties used for property verification were identical to those used by Bruffaerts et al. (2013b) and were the same between the

two experiments. Stimuli and properties were originally derived from a feature applicability matrix collected by De Deyne et al. (2008). Twenty-four animate entities from 6 semantic subcategories (birds, marine animals, fish, herpetofauna, insects, and livestock) were selected based on hierarchical clustering analysis (Figure 1) (Bruffaerts et al., 2013b; Liuzzi et al., 2015). Written word length was between 3 and 11 characters and spoken word duration between 0.33 s and 1.5 s. Word frequency values,which are based on a corpus of 42,380 written words and are a logarithmic function of the lemma counts from the Dutch version of the CELEX database (Baayen et al., 1993), were between 0 and 5.21, age of noun acquisition between 4 and 10.79 years and familiarity between 1.96 and 3.9 (on a 7-point Likert-type scales (De Deyne et al., 2008)). From the feature applicability matrix (De Deyne et al., 2008), a 24-by-24 semantic cossimilarity matrix (Liuzzi et al., 2015) was derived containing the pairwise cosine similarity between each pair of entities (see below).
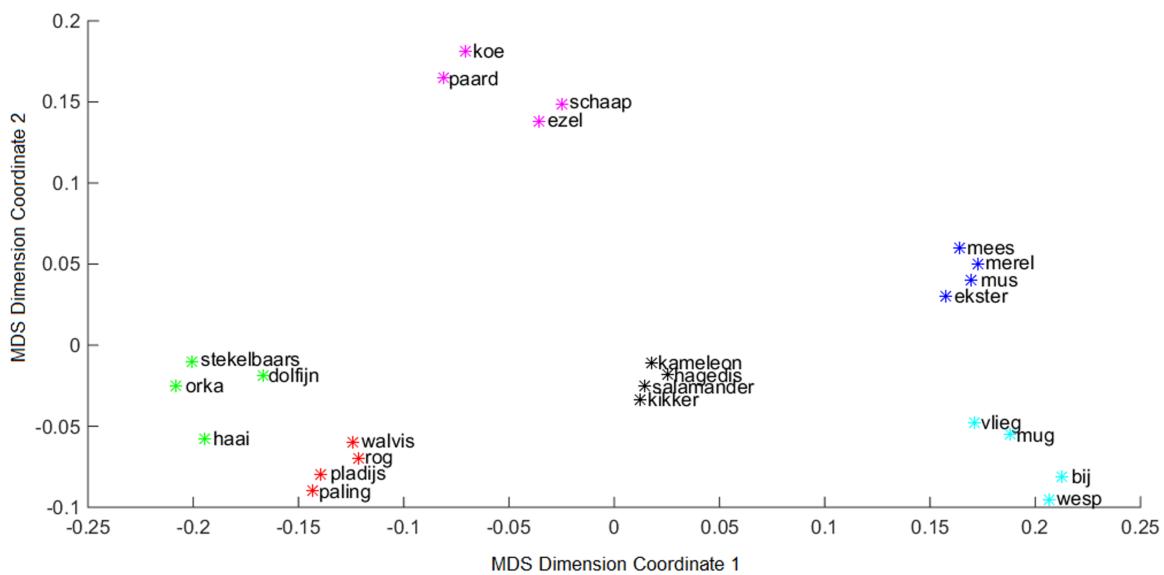

Figure 1: **MDS.** Visual representation of the semantic clusters and semantic distances between entities, based on the feature generation data collected by De Deyne et al. (2008). For visualization, data reduction of the similarity matrix to two dimensions was performed by means of multidimensional scaling (MDS).

The primary analysis of the current study was based on semantic similarities derived from the feature applicability matrix. The results were verified by means of a semantic similarity matrix derived from word associations based on a word association task (De Deyne et al., 2013, 2016). While a semantic similarity matrix based on feature generation may be biased towards sensory and functional-associative features, a semantic similarity matrix based on word associations may provide a more comprehensive estimate of the semantic content of concrete words.

For our set of 24 entities, we also determined the 24-by-24 phonological distance matrix and the orthographic distance matrix. The phonological and orthographic distance

matrices were computed by using the Levenshtein distance (Levenshtein, 1966) between each pair of the 24 words, which corresponds to the minimal number of steps necessary to transform an orthographic or phonological string, respectively, in another one by character substitutions, insertions, or deletions. The phonological transcriptions for the stimuli were obtained from the CELEX lexical database (Baayen et al., 1995). To determine the similarity between the semantic cossimilarity matrix and the phonological and the orthographic distance matrix, we converted the semantic cossimilarity matrix in a semantic cosine distance matrix by subtracting 1 from each cell of the matrix. The Pearson correlation between the semantic cosine distance matrix and the phonological distance matrix was not significant ($P = 0.95$) and neither was this the case for the orthographic distance matrix ($P = 0.86$).

Written words were presented with a letter size of 0.7 visual degrees, spoken words had an average RMS (root mean square) power of -25.33 dB and they were presented with an intensity of +10 dB. All manipulations of the spoken words were performed using Praat (`http://www.fon.hum.uva.nl/praat/`), a scientific tool for analyzing spectrograms.

2.3 Task
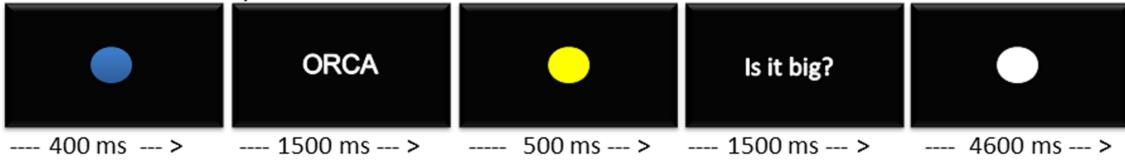
In both experiments, subjects performed the same property verification task (Figure 2 A, B): they were asked to judge whether a given property (the test stimulus) is applicable to a given animal (the sample stimulus). Each property verification trial started with a blue fixation point (duration 400 ms). Next a sample stimulus was presented for 1500 ms, consisting of either a written word in the visual condition or a spoken word in the auditory condition. The sample stimulus was followed by a yellow fixation point (500 ms duration). Next the property verification question was presented for 1500 ms. In the visual condition the property verification question was displayed on the screen, in the auditory condition it was presented auditorily. The questions were presented as fully grammatical questions, e.g. "is it exotic?" (Liuzzi et al., 2015). 1500 ms after the onset of the property verification question, a white fixation point appeared on the screen. Subjects held a response box in their right hand and pressed a lower or upper button for "yes" or "no". This was counterbalanced between subjects. Subjects were told to respond as soon as they had made a decision and to guess if they were uncertain. For each property, the correct response was positive for approximately half of the entities. The eight properties involved in the task ("large", "legs", "wings", "smooth", "exotic", "mammal", "sea" and "flies") were selected from those generated most frequently by the 1003 students participating in De Deyne et al. (2008) (Bruffaerts et al., 2013b; Liuzzi et al., 2015).
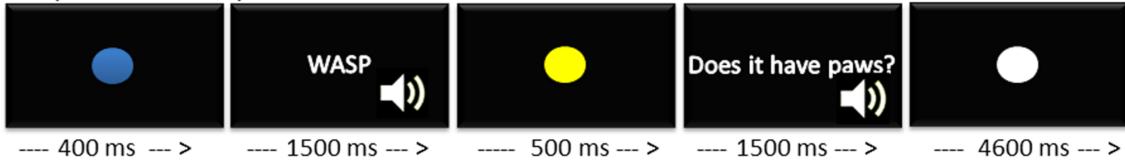
Figure 2: **Task.** **A.** VOI defining experiment: 1. Property verification task in written modality. 2. Property verification task in spoken modality. 3. Low-level control task in written modality. 4. Low-level control task in spoken modality. **B.** Main experiment: 1. Property verification task in written modality. 2. Property verification task in spoken modality. In spoken modality both word and question have been presented in spoken modality.
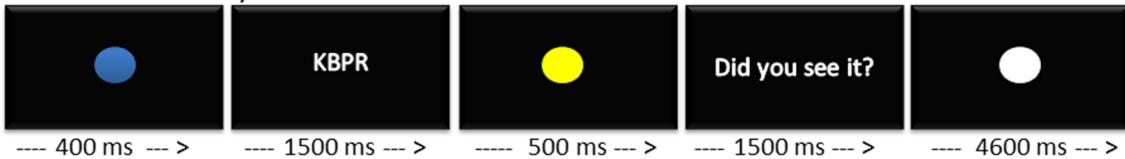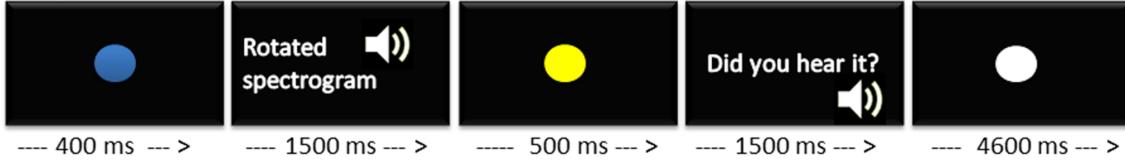
2.4   First experiment

The design of the first experiment was factorial, with two factors: task (two levels: property verification versus control task) and input-modality (two levels: written versus spoken modality) (Figure 2A). The property verification is described above. In the control task, the timing and the sequence of events were the same as in the property verification task. In the control condition for the written words, the sample stimulus consisted of a consonant letter string. The consonant letter string was created by randomizing the position of the letters of the word and replacing the vowels with consonants according to a fixed rule. In the control condition for the spoken words, the sample stimulus consisted of the rotated spectrogram (Scott, 2000). Rotation of the spectrogram abolishes all recognizability while the physical features remain closely matched to the auditory word. In the control conditions, subjects had to respond to a written or spoken question, respectively, asking whether the stimulus was written or auditory. For each modality, half of the questions required a positive answer.

The experiment consisted of 6 runs. Each run (255 scans) was composed of 60 trials, which lasted for 8500 ms each. These trials were composed of 16 property verification trials with written words and 16 with spoken words, 8 control trials with consonant letter strings and 8 with rotated spectrograms and 12 null trials. Null events consisted of a white fixation point which the subject had to fixate. Across all 6 runs each concept appeared 8 times: 4 times as a written word and 4 times as a spoken word. Also, each concept was combined with each property once for each subject. Thus across all 6 runs the combinations stimulus/property were always different. The properties were presented equally often in written as in spoken modality across all subjects.

2.5   MVPA experiment

The second fMRI experiment was optimized for RSA. It contained only two types of trials and more repetitions of a given concept. The two trial types were property verification with written words versus property verification with spoken words. The experiment consisted of 8 runs. In total each concept appeared 16 times across all runs: 8 times as a written word and 8 times as a spoken word. Furthermore, the interval from the end of the probe question till the next trial was 8600 ms instead of 4600 ms so that the haemodynamic response could be extracted without need for deconvolving between rapidly successive trials (Figure 2B).

Each run (300 scans) was composed of 48 trials, which lasted for 12500 ms each. These trials were composed of 24 property verification trials with written words and 24 with spoken words. As in the VOI defining experiment, for each input-modality, every concept was combined with each property once for each subject. Thus across all 8 runs a combination of a given stimulus with a given property never recurred within a specific

input-modality. The properties were presented equally often in written as in spoken modality across all subjects.

2.6 Image acquisition

A 3 tesla Philips Achieva equipped with an 32-channel head volume coil provided functional and structural images. Structural imaging sequences consisted of a T1-weighted 3D turbo-field-echo sequence (repetition time = 9.6 ms, echo time = 4.6 ms, in-plane resolution = 0.97 mm, slice thickness = 1.2 mm). Functional images were obtained using T2* echoplanar images comprising 36 transverse slices (repetition time = 2 s, echo time = 30 ms, voxel size 2.75 x 2.75 x 3.75 mm^3, slice thickness = 3.75 mm, Sensitivity Encoding (SENSE) factor = 2), with the field of view (FOV) (220 x 220 x 135 mm^3) covering the entire brain. Each run was preceded by 4 dummy scans to allow for saturation of the BOLD signal.
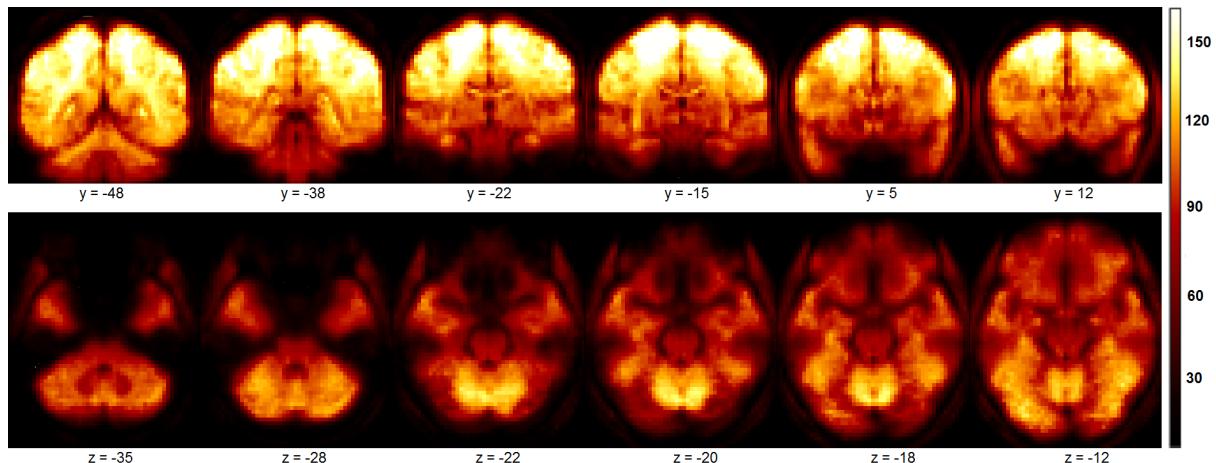

Figure 3: **TSNR.** Temporal signal-to-noise-ratio (TSNR) averaged over all 18 subjects of the univariate experiment. The TSNR is calculated by dividing the mean of the time series by its standard deviation. Task related activity was regressed out.

2.7 Image preprocessing

The images were analyzed with Statistical Parametric Mapping SPM8 (Wellcome Trust Centre for Neuroimaging, University College London, UK). First the data were realigned and resliced and a mean functional image was created. Scans were corrected for slice acquisition time. Next the structural image was co-registered with the mean functional image and segmented in grey matter, white matter and cerebrospinal fluid. Based on the warping parameters obtained during the segmentation step, the functional images were normalized to the Montreal Neurological Institute (MNI) T1 space and resliced to a voxel size of 3 x 3 x 3 mm^3 (Friston et al., 1995). The normalization was also applied to the structural image which was resliced to a voxel size of 1 x 1 x 1 mm^3. Functional images were smoothed using a Gaussian filter with a kernel size of 5 x 5 x 7 mm^3. In

the MVPA experiment, MVPA results were also verified in the absence of smoothing. We used standard SPM8 modelling to remove covariates of no interest (motion regressors, low-frequency trends).

2.8 Behavioral analyses

The reaction times and accuracies from the VOI defining experiment were analyzed by means of a 2-by-2 repeated-measures ANOVA with two within-subjects factors: task (two levels: property verification and control task) and input- modality (two levels: spoken and written words). The reaction times from the MVPA experiment were analyzed by means of a repeated measures ANOVA with input-modality as factor (two-levels: spoken and written words). In both experiments the reaction times were calculated from the onset of the probe question.

Behaviorally, the correlation between semantic similarity and similarities in reaction times was also determined. Semantic similarity between entities may affect the difference in reaction times between entities during the property verification task. To obtain the normalized reaction time difference matrix, the reaction times of each entity were averaged across both modalities, the difference in reaction times between entities was calculated and divided by the sum of reaction times divided by two.

2.9 Univariate analysis of the first experiment

The fMRI data of the VOI defining experiment were modelled using a general linear model (GLM) with 5 event types: property verification trials for written words, control trials for written words, property verification trials for spoken words, control trials for spoken words and null trials. Event onset coincided with the onset of the sample stimulus. The following contrasts were used:

a. Main effect of task: [property verification task with written words + property verification task with spoken words] *minus* [low level control task with written words + low level control task with spoken words])(contrast 1) and the inverse (contrast 2)

b. Main effect of input modality: ([property verification task for written words + low-level control task for written] *minus* [property verification task for spoken words + low level control task for spoken words])(contrast 3) and the inverse (contrast 4)

c. Interaction effect: ([property verification task for written words - low level control task for written words] *minus* [property verification task for spoken words - low level control task for spoken words]) (contrast 5) and the inverse (contrast 6)

The principal contrasts were based on the subtraction of the activation map obtained during the control task from the activation map obtained during the experimental task. In order to determine what was subtracted out, we also performed the following contrast:

d. Control task: ([Low level control task for written words + low level control task for spoken words] *minus* [null events]) (contrast 7)

We also determined the global and the logical conjunction contrasts between property verification versus control task for written and for spoken words. Within the SPM terminology, the global conjunction requires the presence of significant activation in either one or both of the two contrasts, the logical conjunction the presence of significant activation in each of the two contrasts.

e. Global conjunction: [property verification task for written words - low level control task for written words] *and/or* [property verification task for spoken words - low level control task for spoken words] (contrast 8)

f. Logical conjunction [property verification task for written words - low level control task for written words] *and* [property verification task for spoken words - low level control task for spoken words] (contrast 9)

For all contrasts, the significance level was set at a voxel-level inference threshold of uncorrected $P < 0.001$ combined with a cluster-level inference of $P < 0.05$ corrected for the whole brain volume (Poline et al., 1997).

2.10 Representational similarity analysis of the second experiment

The representation similiarity analysis was based on a Pearson correlation analysis between the semantic cossimilarity matrix and the fMRI cossimilarity matrix. In order to obtain the fMRI cossimilarity matrix, for each subject and for each trial, the corrected time series was extracted from the normalized, smoothed fMRI data and used as input from which the fMRI response patterns were obtained by calculating the area under the curve of the BOLD response between 2 and 8 seconds after the start of the sample stimulus within every voxel (Bruffaerts et al., 2013b; Liuzzi et al., 2015). These voxel-based responses were then extracted per VOI. For each pair of trials, the cosine similarity between the fMRI response patterns was calculated and a 384-by-384 matrix (384 being the total number of trials for the active condition) was then obtained. This matrix is called the "result matrix". From this result matrix, four fMRI cossimilarity matrices were derived. The fMRI cossimilarity matrices for written words (24-by-24 matrix, 24 being the total number of entities) was derived by selecting pairs consisting of written words only (e.g. donkey(written)/orca(written)) and likewise for the 24-by-24

fMRI cossimilarity matrix for spoken words (e.g. donkey(spoken)/orca(spoken)). The fMRI cossimilarity matrix for written and spoken words pooled together (24-by-24 matrix) was composed of written-written, spoken-spoken and written and spoken pairs. To derive the 24-by-24 cross-modal fMRI cossimilarity matrix, we selected and averaged the pairs composed of stimuli referring to different entities through different input modalities (e.g. donkey(written)/ orca(spoken) *and* donkey(spoken)/orca(written)). Each pair was averaged within subjects and across subjects.

The semantic cossimilarity matrix was obtained by selecting from the concept-feature matrix the entire columns (*764 rows*) related to the 24 animate entities used in the experiment and calculating the cosine similarity between each pair of columns.

The representational similarity was calculated by means of Pearson regression analysis between the lower-triangle part of the 24-by-24 semantic cossimilarity matrix and the lower-triangle part of the 24-by-24 fMRI cossimilarity matrix. The diagonal of the matrices, consisting of pairs of the same entities, was not included in the RSA analysis so that the effect of semantic similarity between entities would not be confounded by the effect of identity between entities.

The main analysis was based on the semantic similarity matrix based on feature generation. In order to determine the generalizability of our results to other estimates of semantic similarity, the RSA was also performed with the semantic similarity matrix based on word associations.

2.10.1 Primary analysis: Cross-modal semantic similarity effects

Our primary analysis focused on cross-modal pairs (pairs consisting of a written and a spoken word that refer to different concrete entities). Cross-modal pairs consisting of the same entities were not included so as to eliminate any contribution from identity. The primary outcome analysis consisted of the RSA (Kriegeskorte et al., 2008; Bruffaerts et al., 2013b) between the fMRI cossimilarity matrix for cross-modal pairs and the semantic cossimilarity matrix. The RSA was restricted to the VOIs obtained from contrast *1*. The RSA was based on Pearson regression analysis. The significance of results was determined by performing a 10.000 random permutation labelling of the matrix based on fMRI data. By comparing the true correlation with the 10.000 possible results obtained by random labelling, we could determine the significance of the result. We used a one-tailed statistical threshold of $P < 0.05$ (Bruffaerts et al., 2013a) corrected for the number of VOIs obtained in contrast *1* (the main effect of task: active task > control task).

In order to verify the specifically semantic nature of the effect in regions that showed a significant cross-modal semantic similarity effect we also evaluated whether the phonological distance, the number of phonemes, the orthographic distance or the reaction times played a role in such effect. The fMRI cossimilarity matrix was converted into an fMRI dissimilarity matrix by computing 1-cosine similarity. A representational similarity anal-

ysis was conducted between the fMRI dissimilarity matrix and the dissimilarity matrix for phonological distance, orthographic distance, number of phonemes and reaction times differences, respectively.

2.10.2 Secondary analysis: input-modality specific semantic similarity effects

A secondary analysis was performed within all VOIs obtained in contrasts *3-6* and focused on within-modality pairs (written only and spoken only, respectively). The significance threshold was set at a one-tailed statistical threshold of $P < 0.05$ (Bruffaerts et al., 2013a) corrected for the number of VOIs obtained in contrasts *3-6*.

3 Results

3.1 Behavioral analysis

In the VOI defining experiment, the main effect of task on reaction times was significant ($F(1,17) = 36.147$; $P = 0.00001$) as was the main effect of input-modality ($F(1,17)= 223.68$; $P = 0,00000$), without interaction effect between the two factors ($F(1,17) = 0.00046$; $P = 0.98315$). Responses were significantly slower during the experimental than during the control condition and during the auditory than during the written conditions (Table 1). The effect of input modality on reaction times was confirmed in the second experiment ($F(1,19) = 117.58$, $P = 0.00000$) and accuracies were also similar to those obtained in the first experiment (mean 93.15%, S.D. 1.32).

Reaction time and Accuracy of response

	VOI defining experiment							
	Written		Control written		Spoken		Control spoken	
	Mean	S.D	Mean	S.D	Mean	S.D	Mean	S.D
Reaction Time	1.63 s	0.47	1.45 s	0.45	2.16 s	0.39	1.98 s	0.34
Accuracy	93.6%		96.5%		93.25%		96.1%	
	MVPA experiment							
	Written				Spoken			
	Mean	S.D			Mean	S.D		
Reaction Time	1.76 s	0.44			2.24 s	0.42		
Accuracy	92.22%				94.09%			

Table 1: Reaction times and accuracies of responses for the VOI-defining experiment and for the MVPA experiment. Reaction times were calculated from the start of the probe question onwards. *Abbreviations:* S.D.: standard deviation.

Reaction time differences between entities during property verification correlated significantly with the semantic distances derived from feature generation (Pearson regression coefficient = 0.24 , $P = 0.001$) or word association (Pearson regression coefficient = 0.30, $P = 0.0001$).

3.2 First experiment

Compared to the control condition, the property verification task activated the retrosplenial cortex, the left ventral occipitotemporal transition zone (vOT) and left ventromedial temporal cortex (encompassing the left perirhinal cortex (Bruffaerts et al., 2013b; Liuzzi et al., 2015)), the anterodorsal part of the left pars triangularis, the pars orbitalis bilaterally, and the frontal pole (contrast *1*, main effect of task; Table 2; Figure 4A).

The inverse contrast (contrast *2*) yielded activation of the angular gyrus bilaterally (Figure 4B and Figure 7A) and the precuneus (Figure 4B), among other regions.

	MNI coordinates			Extent	$P_{\text{FWE-corr.}}$ (cluster-level)
	x	y	z		
Retrosplenial cortex	-3	-55	7	882	0.000
	-12	-58	-11		
	-9	-46	-2		
LIFG, orbital part	-30	35	-14	99	0.000
RIFG, orbital part	33	35	-14	40	0.017
	30	41	-8		
LIFG, triangular part	-39	26	13	53	0.004
	-45	35	13		
Frontal pole	0	59	-8	33	0.038
L vOT	-51	-55	-14	68	0.001
	-48	-43	-17		
L ventromedial temporal cortex	-24	-22	-17	317	0.000
	-36	-13	-23		
	-33	-34	-20		

Table 2: First experiment: Main effect of task where the property verification yields higher activity than the control condition. Clusters showing a main effect of task (contrast *1*) and clusters obtained from the conjunction analysis (contrast *3*) at a voxel-level inference threshold of uncorrected $P < 0.001$ combined with a cluster-level inference of $P < 0.05$ corrected for the whole brain volume (Poline et al., 1997). Extent refers to the number of 3 x 3 x 3 mm^3 voxels. *Abbreviations:* L: left; LIFG: left inferior frontal gyrus; RIFG: right inferior frontal gyrus; vOT: ventral occipitotemporal transition zone.
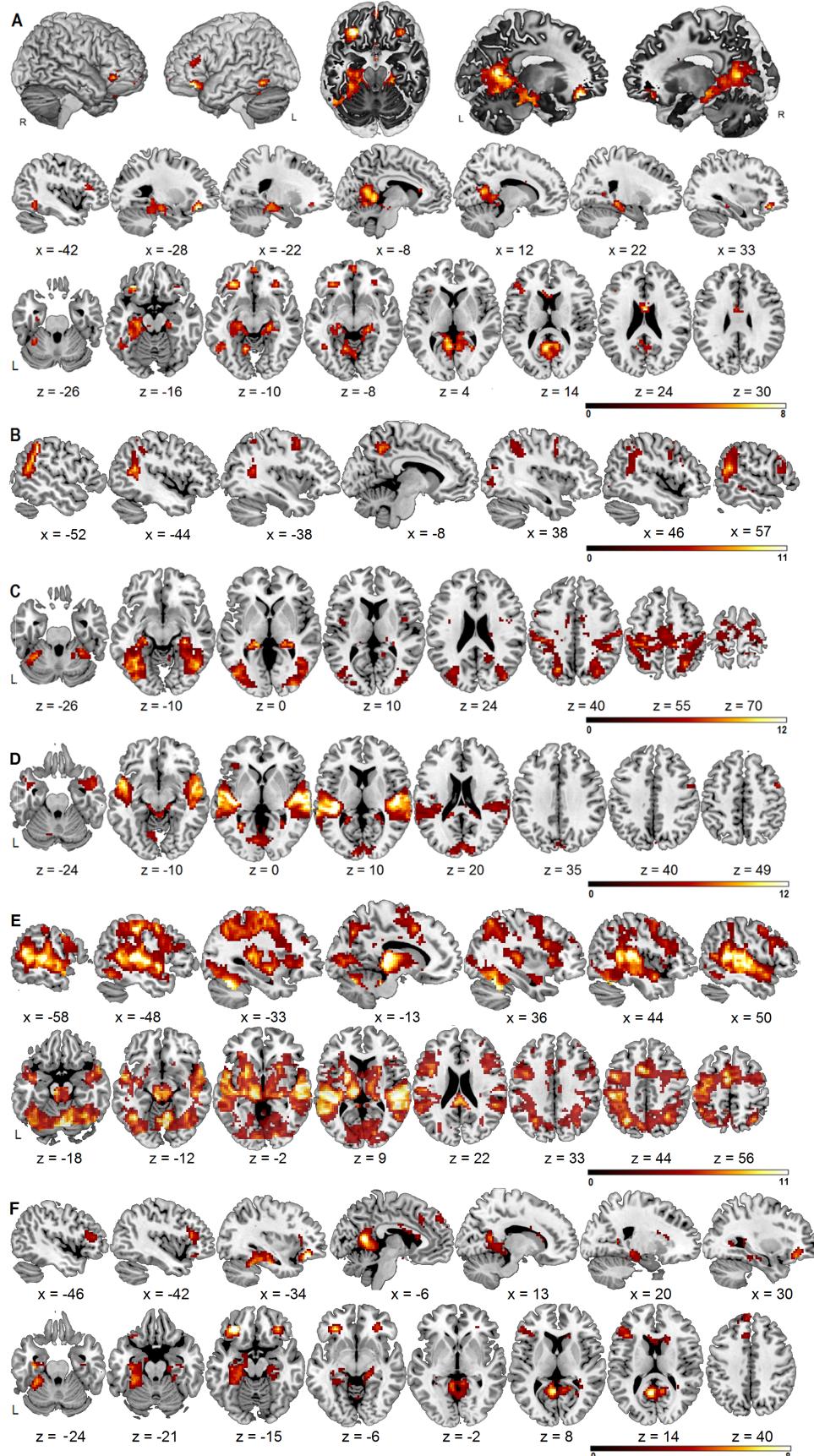
Figure 4: **Univariate contrasts. A.** Main effect of task: active task > control task (contrast *1*): 3D rendering, sagittal and axial slices. **B.** Main effect of task: control task > active task (contrast *2*): sagittal slices. **C.** Main effect of input modality: visual > auditory modality (contrast *3*): axial slices. **D.** Main effect of input modality: auditory > visual modality (condition *4*): axial slices. **E.** Control task > null baseline (contrast *7*): sagittal and axial slices. **F.** Global Conjunction (contrast *8*): sagittal and axial slices of the first experiment at a voxel level inference threshold of uncorrected $P < 0.001$ combined with a cluster-level inference of $P < 0.05$ corrected for the whole brain volume (Poline et al., 1997).

The global conjunction analysis (contrast *8*) yielded a pattern very similar to the main effect of task, including the pars triangularis among other regions (Figure 4F, Supplementary Table 2). The logical conjunction analysis (contrast *9*) yielded significant activation of the pars orbitalis of the LIFG, retrosplenial cortex and left ventromedial temporal cortex (Supplementary Table 2). In the logical conjunction analysis, the triangular part was activated in both contrasts at voxel-level inference threshold of uncorrected $P < 0.001$ ($Z = 3.92$, $P_{uncorr.} = 0.000$) but the extent of this conjoint activation did not reach the cluster-level corrected significance threshold. The time activity curve derived from contrast *1* convincingly showed activation both for written and spoken words compared to the respective control conditions. The results for the main effect of input modality (contrast *3* and *4*) are shown in Figure 4C and Figure 4D (Supplementary Table 1). For the interaction effects, inspection of the event-related time-activity curves indicates that in most instances the interaction was due to a differential decrease in the active compared to the control condition, with the highest activity in the visual control condition. In left vOT, activity was highest in both visual conditions and also in the spoken word condition compared to the auditory control condition (Supplementary Figure 1).

Finally, compared to the null baseline, the control conditions activated a large set of regions shown in Figure 4E. Notably, the control condition did not activate the posterior or anterior middle and inferior temporal gyrus so that the absence of activation in these regions in contrast *1* cannot be accounted for by their activation during the control condition.

3.3 Primary outcome analysis: cross-modal semantic similarity effect

The primary outcome analysis of this study consisted of the RSA between the fMRI cossimilarity matrix for cross-modal pairs and the semantic cossimilarity matrix, in those regions activated in univariate contrast *1*.

Each of the 7 VOIs obtained from contrast *1* (Table 2) was subjected to an RSA for cross-modal semantic similarity (word pairs consisting exclusively of one written and one spoken word referring to different concrete entities). Among these regions, only the left anterodorsal pars triangularis showed a cross-modal semantic similarity effect (Pearson correlation (r) = 0.20, $P_{corr.} = 0.0014$): activity patterns were more similar for word pairs that were more semantically similar, even though the words were presented in two different modalities (Figure 5).

The semantic similarity effect was also significant in this region when we pooled all pairs regardless of input-modality (r = 0.21, $P_{corr.} = 0.0028$). When the semantic similarity matrix was based on word associations rather than feature generation, the RSA between the semantic cossimilarity matrix based on word associations and the fMRI cossimilarity matrix for crossmodal pairs in the left anterodorsal pars triangularis showed a significant semantic similarity effect (r = 0.13, $P = 0.01$).

The similarity effect obtained in the left pars triangularis was specifically semantic: There was no correlation with the phonological distance or orthographic distance matrix ($P > 0.8$). Neither there was any correlation with the difference in number of phonemes ($P > 0.9$).

In order to examine the role of reaction time differences between entities on the activity pattern in the left pars triangularis, an RSA was conducted between the fMRI dissimilarity matrix and dissimilarity matrix based on reaction times: The correlation was far from

significant ($P > 0.5$). The RSA was also performed at the individual level in order to check the distribution of the individual's correlation values for normality. The distribution of the correlation values did not significantly deviate from normality (Shapiro-Wilk test, $P = 0.17$) and differed significantly from the null distribution (Students t test, $P = 0.037$). The distribution of the Z scores did not significantly deviate from normality (Shapiro-Wilk test, $P = 0.98$) and differed significantly from the null distribution (Students t test, $P = 0.036$).
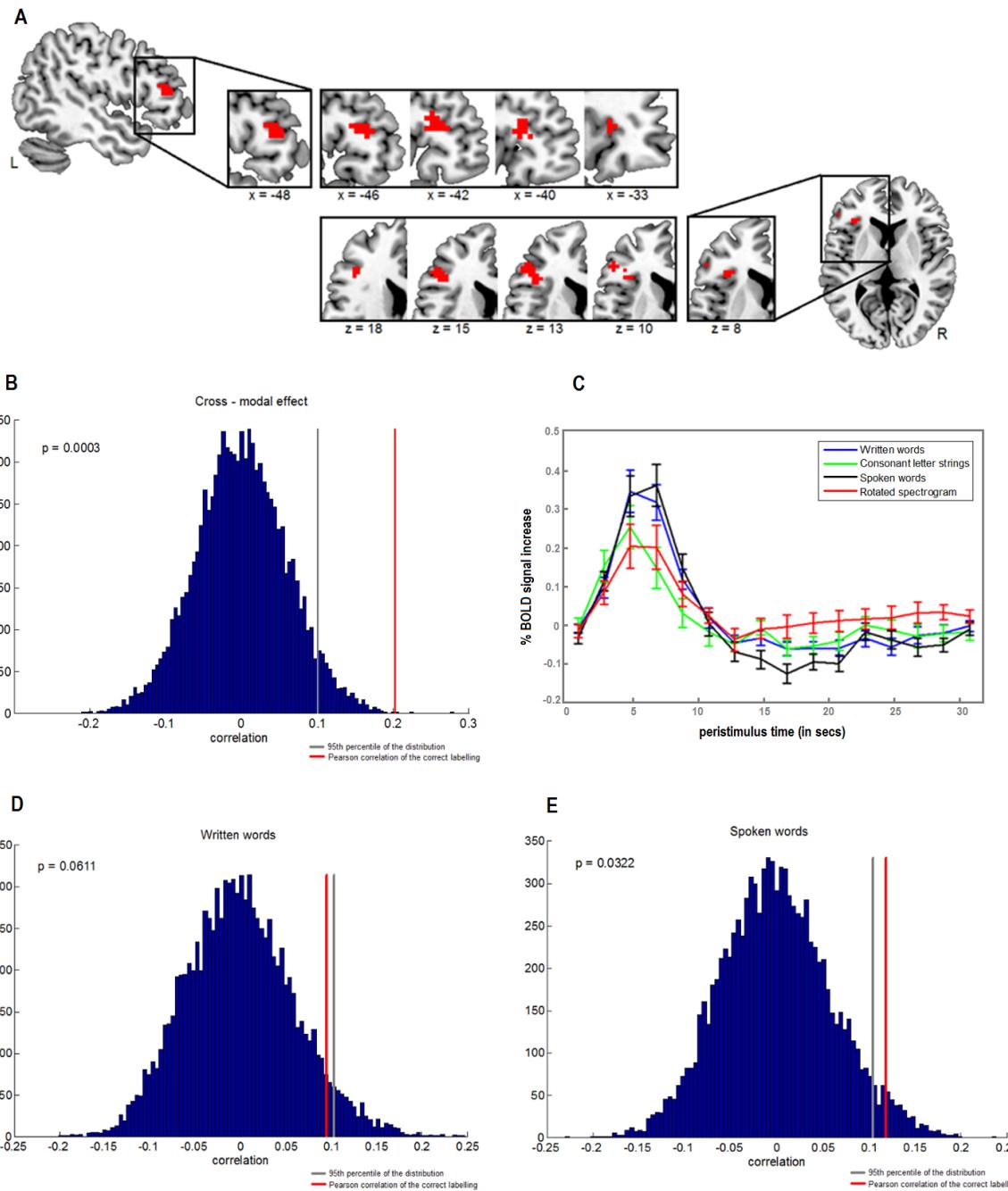
Figure 5: **Cross-modal semantic similarity effect in left anterodorsal pars triangularis**. **A.** Sagittal and axial slices of the left anterodorsal pars triangularis as binary VOI. **B.** Probability distributions for the representational similarity analysis (RSA) between the semantic cossimilarity matrix and the fMRI cossimilarity matrix for cross-modal effect after random labelling. **C.** Plots of the peristimulus response function, based on VOI defining experiment dataset, showing the percent signal change of property verification trials (blue and black) and control trials (green and red). **D.** Probability distributions for the RSA between the semantic cossimilarity matrix and the fMRI cossimilarity matrix for written words after random labelling. **E.** Probability distributions for the RSA between the semantic cossimilarity matrix and the fMRI cossimilarity matrix for spoken words after random labelling. The red line indicates the cosine similarity between the similarity matrix based on behavioural data (De Deyne et al., 2008) and the similarity matrix based on the fMRI data derived from the response patterns within the VOI. The grey line indicates the 95th percentile of the distribution. X axis: correlation averaged over the group of subjects. Y axis: absolute frequency of a given cosine similarity value across a total of 10.000 random permutation labellings.

Similar effects of crossmodal semantic similarity were obtained when the analysis was based on normalized images without smoothing (r = 0.15, P = 0.007).

In order to determine whether our result depended on the exact way in which we defined the VOIs, a similar analysis was conducted using the cytoarchitectonically defined BA 45 and 44. The definition was based on the Jülich-Düsseldorf cytoarchitectonic atlas, using the Anatomy Toolbox (`http://www.fz-juelich.de/inm/inm-1/DE/Forschung/_docs/SPMAnatomyToolbox/SPMAnatomyToolbox_node.html`) (Eickhoff et al., 2005, 2006, 2007). Brodmann area 45 showed a semantic similarity effect for cross-modal pairs (r = 0.15, $P_{corr.}$ = 0.017), for written words (r = 0.12, $P_{corr.}$ = 0.04), and for written and spoken words pooled together (r = 0.19, $P_{corr.}$ = 0.002). Results were Bonferroni-corrected for number for VOIs. No semantic similarity effects were detected in cytoarchitectonically defined Brodmann area 44: semantic similarity effect for written words (r = -0.006, $P_{uncorr.}$ = 0.52), written and spoken words pooled together (r = 0.02, $P_{uncorr.}$ = 0.35) and for cross-modal pairs (r = 0.02, $P_{uncorr.}$ = 0.36).

In BA45, the phonological nor the orthographic distance matrix correlated with the fMRI dissimilarity matrix for written words (P = 0.9 for phonological and P = 0.8 for orthographic effect). Neither was there any correlation with the fMRI dissimilarity matrix for spoken words (P = 0.3 for phonological and P = 0.4 for orthographic effect), nor with the fMRI dissimilarity matrix for cross-modal pairs (P = 0.5 for phonological and P = 0.4 for orthographic effect). In contrast, in BA44 the orthographic (r = 0.1433 P = 0.05) and the phonological (r = 0.1385 P = 0.06) similarity effect tended to correlate with the fMRI similarity matrix for spoken words. No correlation was found for written words nor cross-modally.

The volumes we tested differed in size and, theoretically, size may affect the sensitivity of MVPA, hence we evaluated whether we could find cross-modal similarity effects in other VOIs if size was shrunk to 53 voxels. Retrosplenial cortex (882 voxels) contained 16 local maxima in the univariate analysis of the first experiment. We defined VOIs of 53 voxels in size centred around these local maxima. In none of these VOIs was there a semantic similarity effect at uncorrected P > 0.05. A similar procedure was applied in left ventromedial temporal cortex where the univariate contrast revealed 13 local maxima. No cross-modal semantic similarity effects were detected in these VOIs at uncorrected P > 0.05.

For the sake of completeness we also determined semantic similarity effects for within-modality pairs in the regions obtained from contrast *1* (Supplementary Table 3). In the left anterodorsal pars triangularis a trend was present for within-modality pairs for the spoken and for the written modality (Figure 5). Similar results for written (r = 0.11, P = 0.03) and spoken (r = 0.10, P = 0.04) modality were obtained when we performed the RSA by using the semantic cossimilarity matrix based on word associations (De Deyne et al., 2013, 2016). In line with previous findings (Bruffaerts et al., 2013b; Liuzzi et al., 2015), the left ventromedial temporal cortex showed a semantic similarity effect for written word pairs (r = 0.15 $P_{uncorr.}$ = 0.009) (Figure 6).
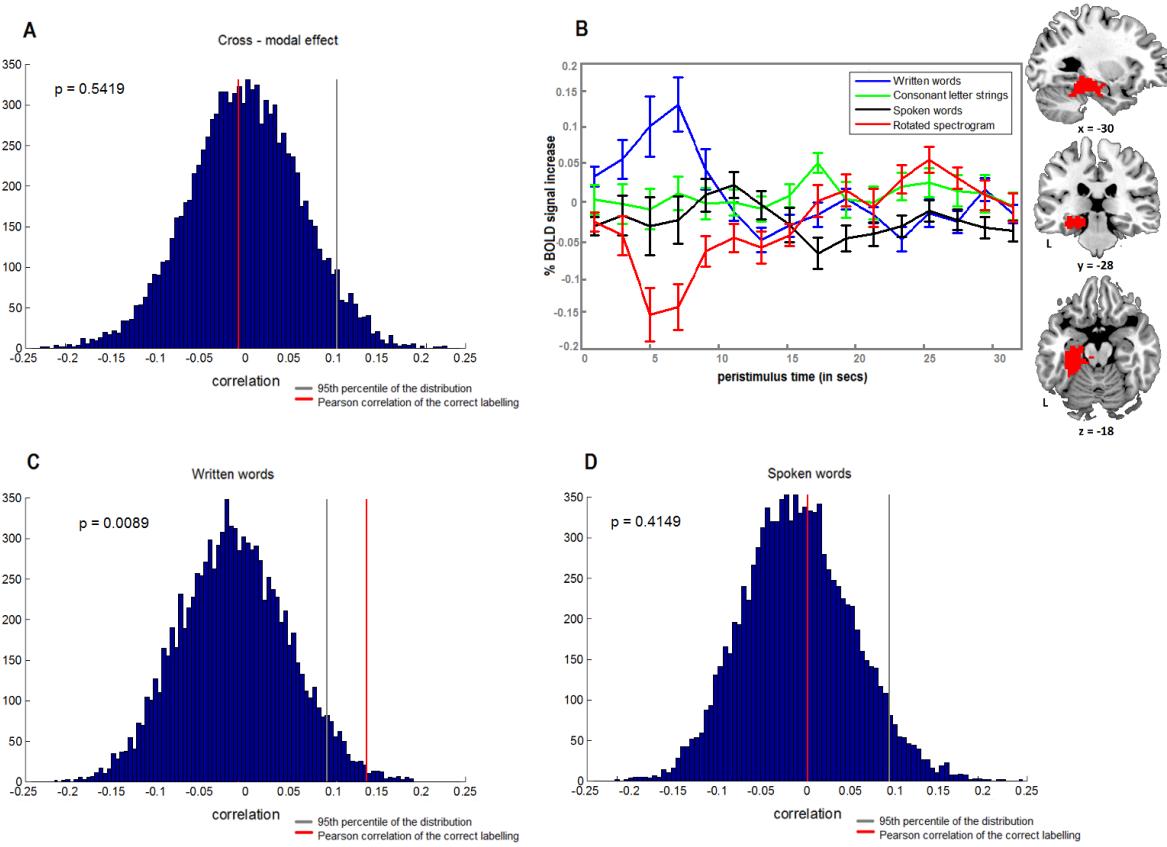
Figure 6: **Semantic similarity effect for written words in left ventromedial temporal cortex**. **A.** Probability distributions for the RSA between the semantic cossimilarity matrix and the fMRI cossimilarity matrix for cross-modal effect after random labelling. **B.** Sagittal, coronal and axial slice of the left ventromedial temporal cortex and plots of the peristimulus response function, based on VOI defining experiment dataset, showing the percent signal change of property verification trials (blue and black) and control trials (green and red). **C.** Probability distributions for the RSA between the semantic cossimilarity matrix and the fMRI cossimilarity matrix for written words after random labelling. **D.** Probability distributions for the RSA between the semantic cossimilarity matrix and the fMRI cossimilarity matrix for spoken words after random labelling.

The left angular gyrus was more active during the control condition than during the property verification. Since the angular gyrus has also been implicated in semantic processing (Binder et al., 2009), an exploratory analysis was performed where an RSA was conducted between the cross-modal fMRI activity similarity matrix in the left angular gyrus and the semantic similarity matrix. No cross-modal semantic similarity effect were detected with the semantic similarity based on concept-feature matrix (r = 0.07, p = 0.1368) nor with that based on word-association (r = 0.04, p = 0.23) (Figure 7).
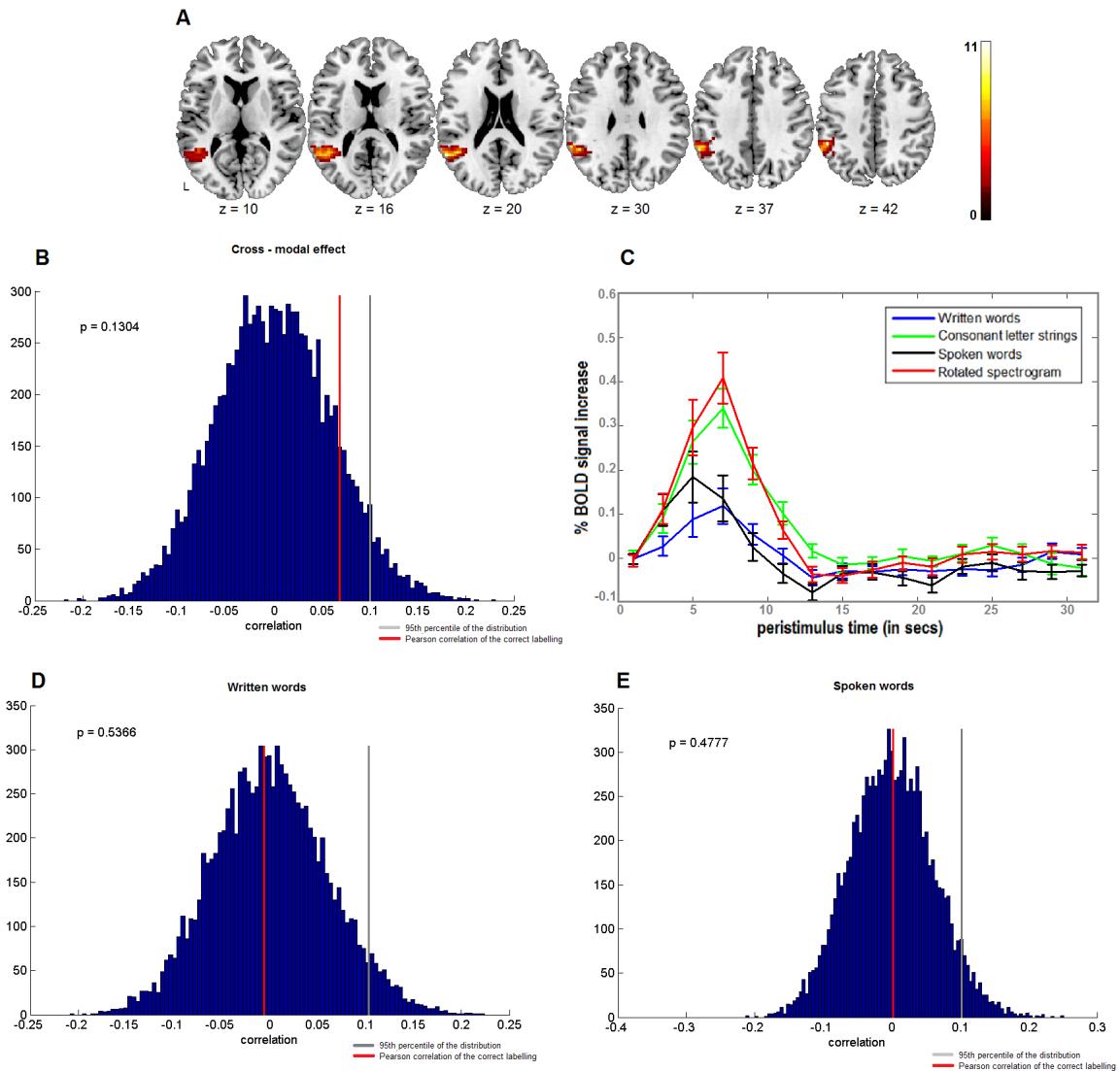
Figure 7: **Angular Gyrus. A.** Axial slices of the Angular Gyrus based on main effect of task (contrast 2). **B.** Probability distributions for the representational similarity analysis (RSA) between the semantic cossimilarity matrix and the fMRI cossimilarity matrix for cross-modal effect after random labelling. **C.** Plots of the peristimulus response function, based on VOI defining experiment dataset, showing the percent signal change of property verification trials (blue and black) and control trials (green and red). **D.** Probability distributions for the RSA between the semantic cossimilarity matrix and the fMRI cossimilarity matrix for written words after random labelling. **E.** Probability distributions for the RSA between the semantic cossimilarity matrix and the fMRI cossimilarity matrix for spoken words after random labelling.

3.4 Secondary analysis: Within-modality semantic similarity effects

A secondary outcome analysis of this study aimed at localizing semantic similarity effects within a given input modality. This was restricted to those clusters that were derived from the main effect of input modality and from the interaction effect (contrasts 3-6) in the first experiment (total number of VOIs = 18).

The left STG exhibited a semantic similarity effect for spoken words (r = 0.17, $P_{corr.}$ = 0.04) (Table 3; Figure 8). The semantic similarity effect for spoken words in the left STG

was further investigated by partitioning this VOI in five equally sized partitions along the posterior-anterior direction perpendicularly to the superior temporal sulcus (Figure 8E). The partitions 2 (r = 0.16, $P_{\text{corr.}}$ = 0.04), 3 (r = 0.18, $P_{\text{corr.}}$ = 0.0085) and 4 (r = 0.20, $P_{\text{corr.}}$ = 0.0035) showed a semantic similarity effect for spoken words (Table 4; Figure 8E).

RSA - VOIs based on the main effect of auditory modality

		r	P_{uncorr}
Left STG	Cross-modal effect	-0.11	0.96
	Written and spoken words pooled	0.11	0.04
	Written words only	-0.009	0.54
	Spoken words only	0.17	0.002 *
Right STG	Cross-modal effect	-0.10	0.94
	Written and spoken words pooled	0.14	0.01
	Written words only	0.03	0.33
	Spoken words only	0.17	0.003

RSA - VOIs based on the interaction effect

		r	P_{uncorr}
Left vOT	Cross-modal effect	-0.04	0.73
	Written and spoken words pooled	0.11	0.04
	Written words only	0.18	0.004
	Spoken words only	0.02	0.35

Table 3: MVPA experiment. RSA of activity patterns in VOIs defined by means of contrasts 5 and 7 of the first experiment. Values which reached significance after correction for the number of VOIs tested are marked with an asterisk (number of VOIs = 18, $P_{\text{corr.}} < 0.05$). *Abbreviations:* STG = superior temporal gyrus; vOT: ventral occipitotemporal transition zone.

Superior Temporal Gyrus split in 5 partitions

		r	P_{uncorr}
Left STG - Partition 1	Cross-modal effect	-0.06	0.92
	Written and spoken words pooled	0.11	0.05
	Written words only	0.004	0.46
	Spoken words only	0.14	0.02
Left STG - Partition 2	Cross-modal effect	-0.1	0.93
	Written and spoken words pooled	0.12	0.03
	Written words only	0.003	0.48
	Spoken words only	0.16	0.0072 *
Left STG - Partition 3	Cross-modal effect	-0.14	0.98
	Written and spoken words pooled	0.12	0.03
	Written words only	0.06	0.19
	Spoken words only	0.18	0.0017 *
Left STG - Partition 4	Cross-modal effect	-0.08	0.9
	Written and spoken words pooled	0.1	0.06
	Written words only	-0.11	0.96
	Spoken words only	0.2	0.0007 *
Left STG - Partition 5	Cross-modal effect	-0.04	0.75
	Written and spoken words pooled	0.04	0.25
	Written words only	-0.05	0.76
	Spoken words only	0.14	0.02

Table 4: MVPA experiment. RSA based on the 5 partitions of the left STG. From 1 to 5, the partitions move progressively to the more anterior portion of the left STG along the posterior-anterior axis. Values which reached significance after correction for the number of VOIs tested are marked with an asterisks (number of partitions = 5, $P_{\text{corr.}} < 0.05$). *Abbreviations:* STG = superior temporal gyrus.
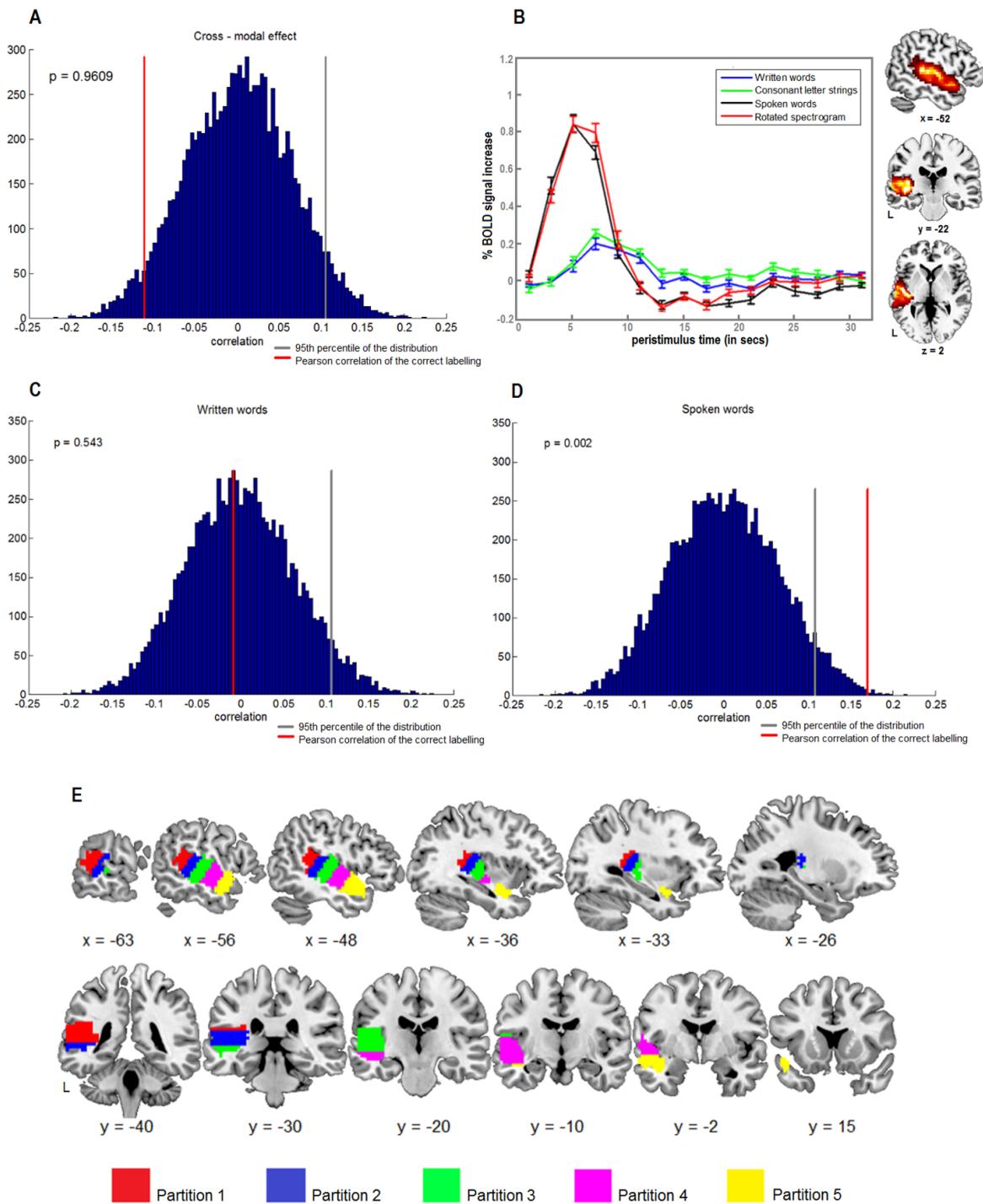
Figure 8: **Semantic similarity effect for spoken words in left superior temporal gyrus**. **A.** Probability distributions for the RSA between the semantic cossimilarity matrix and the fMRI cossimilarity matrix for cross-modal effect after random labelling. **B.** Sagittal, coronal and axial slice of the left superior temporal sulcus-gyrus and plots of the peristimulus response function, based on VOI defining experiment dataset, showing the percent signal change of property verification trials (blue and black) and control trials (green and red). **C.** Probability distributions for the RSA between the semantic cossimilarity matrix and the fMRI cossimilarity matrix for written words after random labelling. **D.** Probability distributions for the RSA between the semantic cossimilarity matrix and the fMRI cossimilarity matrix for spoken words after random labelling. **E.** Sagittal and axial slices of the binary left superior temporal gyrus divided in 5 partitions along the posterior-anterior direction.

4 Discussion

Previous fMRI studies using univariate analysis (Vandenberghe et al., 1996; Chee et al., 1999; Wagner et al., 2001; Booth et al., 2002; Homae, 2002; Constable et al., 2004; Gold et al., 2005a; Spitsyna et al., 2006) have implicated BA45 in semantic processing. In the current study, using RSA we demonstrate that BA45 codes for the semantic relationships between entities and makes abstraction of the exact input modality in which these entities were presented. This novel finding provides insight in the exact cognitive operations carried out in BA45 during semantic tasks and has consequences for where we place the triangular part in relation to other regions which are involved in coding meaning such as the anterior superior temporal and the ventromedial occipitotemporal cortex.

The cross-modal effect in the left pars triangularis does not relate to shared phonological retrieval for written and spoken words: The semantic cossimilarity matrix did not show any correspondence with the orthographic or phonological similarity matrix. Furthermore, in the pars triangularis the phonological and the orthographic similarity matrix did not correlate with the fMRI cossimilarity matrix. Neither did the difference in number of phonemes correlate with either the semantic similarity matrix or the fMRI cossimilarity matrix. Previous studies implicated posterior inferior frontal gyrus rather than the triangular part in phonological processing (Wagner, 2000; Wagner et al., 2001; McDermott et al., 2003; Devlin et al., 2003; Katzev et al., 2013).

The volumes-of-interest were defined for further MVPA based on a univariate contrast between the property verification task and a low-level control condition from an independent experiment. In the control condition the sample stimuli consisted of consonant letter strings or auditory stimuli obtained from rotation of the spectrogram as well as a probe question. We did not use pseudowords as sample stimuli as they may elicit semantic activation (Price et al., 1996). The probe question was chosen to be simple and have a minimal semantic control load. Compared to the active condition, the control condition activated the default mode network, including the angular gyrus among others (Figure 4B and Figure 7A). Compared to baseline, the control condition did not activate the associative semantic network (included regions such as the middle temporal gyrus) to any significant degree due to its very low cognitive load.

Semantic similarity between entities correlated with reaction times during property verification. The pars triangularis lies relatively close to the inferior frontal sulcus which has been implicated in cognitive control (Duncan and Owen, 2000). There was no correlation between the reaction time differences between entities and the fMRI cossimilarity matrix in the left triangular part, rendering the cognitive control an unlikely explanation of the semantic similarity effect.

A cornerstone of the RSA in the current study is the semantic similarity matrix. The way it is constructed will critically determine the outcome. The main analysis was based on a semantic similarity matrix derived from a feature generation task. This could have biased the results towards a similarity based on sensory and functional-associative features. The outcome was verified by means of an RSA with a semantic similarity based on word associations (De Deyne et al., 2013, 2016). By reflecting the imagery, knowledge, beliefs, attitudes, and affect in the associations, a similarity matrix based on word associations may quantify the mental representation in a more comprehensive way (De Deyne et al., 2015). Results were entirely comparable.

The activity focus in the pars triangularis lay near the lower bank of the inferior frontal sulcus (IFS) (Uylings et al., 1999) (Figure 5), in or near Brodmann's area (BA)

45A (Uylings et al., 1999; Friederici, 2002; Petrides and Pandya, 2009). The crossmodal semantic similarity effect was confirmed when the inferior frontal region was defined in a different way, i.e. based on a cytoarchitectonic atlas (Eickhoff et al., 2005, 2006, 2007). The anatomical connectivity of BA45A in nonhuman primates differs from that of BA47/12. Whereas BA47/12 is principally connected with inferotemporal and perirhinal cortex (Petrides and Pandya, 2002; Saleem et al., 2008; Gerbella et al., 2010), BA45A also has dense connections with auditory parabelt cortex, along with visual association cortex and the superior temporal polysensory (STP) cortex (Gerbella et al., 2010). The main pathway through which BA45A is connected with auditory belt cortex is the extreme capsule (Frey et al., 2008; Petrides and Pandya, 2009; Friederici and Gierhan, 2013). Connectivity with auditory belt areas is relatively unique for BA45A among the inferior frontal regions (Gerbella et al., 2010) and distinct from e.g. BA45B. Although one has to be careful drawing interspecies homologies, the cross-modal semantic similarity effect in pars triangularis seems to recapitulate the anatomical connectivity of BA45A with both visual and auditory association cortex.

Standard probabilistic cytoarchitectonic mapping of area 45 and area 44 (Amunts et al., 1999, 2010) has been used in previous studies. Both areas are more active for pseudowords than for words, with higher activity during lexical than during phonological decision in area 45 but not in area 44 (Heim et al., 2005). Area 45 is more active during semantic generation tasks (such as animal verbal fluency) than during retrieval of over-learned word sequences (Amunts et al., 2004) or letter fluency (Katzev et al., 2013) while area 44 shows the inverse pattern. The differential effect we found in the current study between semantic and phonological similarity of written and spoken words in area 45 and area 44, respectively, is in line with the hypothesis that the pars opercularis is more involved in phonological operations and the anterodorsal part of the pars triangularis in processing meaning (Poldrack et al., 1999).

Our findings indicate that semantic content is, at least momentarily, represented in the pars triangularis. While the data presented cannot rule out that this coding of word meaning is context-independent, it seems much more likely based on prior evidence that the coding of meaning in the triangular part is context-dependent: According to the adaptive coding model of prefrontal cortex (Duncan, 2001), neuronal populations in prefrontal cortex are not inherently tuned to specific features in the world but adapt their tuning profiles to represent input depending on task demands (Stokes et al., 2013). Theories of the organisation of the language system (Hickok and Poeppel, 2007; Lau et al., 2008; Friederici, 2011; Lambon Ralph et al., 2016) differ in the exact function attributed to area 45: semantic working memory (Gabrieli et al., 1998), semantic control (Lambon Ralph et al., 2016), selection (Thompson-Schill et al., 1997), or unification in the sense of integration between word and world knowledge (Hagoort, 2005). The current experiments were not designed with the purpose of adjudicating between these theories. In some models (e.g. the semantic working memory model (Gabrieli et al., 1998)) a representation of meaning is an explicit and integral part of the model, in line with classical working memory theories of prefrontal function (Goldman-Rakic, 1987). Other models do not make strong predictions in this respect to our understanding: the selection (Thompson-Schill et al., 1997) or the semantic control model (Lambon Ralph et al., 2016) could be formulated in a way that there is no representation of semantic similarity within area 45, but certainly leave room for such semantic coding in area 45. This would depend on how interactive the coding of semantic content and the operations 'selection' or 'control' are and how these functions are operationalized in terms of neuronal responses within area 45

and connecting areas. The classical dual stream language and speech model (Hickok and Poeppel, 2007; Lau et al., 2008) is to the best of our knowledge relatively vague about the degree to which the projection zone of the ventral stream into the inferior frontal cortex represents word meaning (Weiller et al., 2011). The current empirical findings convincingly show such a representation in area 45 for both auditory and visual modality. This extension of the ventral word comprehension pathway into area 45 for written and spoken words fits with functional and anatomical connectivity studies between lateral temporal and inferior frontal cortex during auditory language comprehension (Saur et al., 2008, 2010) and extends this to the written word modality. Our novel findings would be incompatible with models that assume that the operations in pars triangularis are carried out on representations of meaning that are present at a distance from pars triangularis, e.g. in posterior temporal cortex. Indeed, such models would fail to explain why semantic similarity affects the response pattern in pars triangularis. The representation of meaning must, at least temporarily, be uploaded in the pars triangularis. Further experiments are needed to evaluate to which degree this effect depends on the need to integrate the word meaning into the broader task context (probe stimulus, verification question). The integration of meaning between the concrete noun and the probe question may have played a role in recruiting the anterodorsal triangular part (Goucha and Friederici, 2015). Apart from the grammatical structure induced by the probe question, the task itself may also have contributed to the observed effect (Wagner, 2000). Hence, both the item processed and the specific task performed may have contributed to the positive effect.

Previous cross-modal classification studies (Simanova et al., 2014) have revealed that semantic representations have a widely distributed nature and previous MVPA studies using written words have demonstrated that distributed modality-specific pathways can be activated as a function of a linear combination of the attributes they code for and that together constitute the meaning of the word (Fernandino et al., 2015). How can the effects we describe be reconciled with distributed neuroanatomical models of semantic processing of attributes and concepts (Binder and Desai, 2011; Huth et al., 2012; Handjaras et al., 2016; Huth et al., 2016)? In line with a semantic working memory account of left inferior frontal cortex (Demb et al., 1995) we speculate that the convergence of spoken and written word input during meaning extraction in pars triangularis serves as a gateway to the distributed processing of meaning in the brain. Once input has converged onto the pars triangularis and gained access to working memory and reaches consciousness (Baars, 1993, 2002; Bundesen et al., 2005; Van Doren et al., 2010a), it is in a suitable position to connect with distributed systems representing the meaning of concepts (Huth et al., 2016). When written words or pictures are identified consciously compared to subliminal processing, a same or nearby inferior frontal region (-42, 36, 15) is activated (Van Doren et al., 2010b). Activation in this region predicts successful memory retrieval (Van Doren et al., 2010b). Until now there was no direct evidence that semantic knowledge was coded in this region. Our data provide for the first time evidence that semantic relationships are coded in the pars triangularis. Researchwise, it remains to be empirically tested whether this distributed activity of feature-specific cortex elicited by concrete nouns precedes or follows the semantic working memory stage in left pars triangularis.

At the downstream ends of the modality-specific pathways, effects were found of semantic similarity within-modality, in line with previous studies (Scott, 2000; Bruffaerts et al., 2013b; Liuzzi et al., 2015). This suggests that associations and conceptual relationships between words are already coded in the visual (Figure 6) and the auditory processing pathways (Figure 8), respectively. A semantic similarity effect for written words is present

in left perirhinal cortex, which belongs to the ventromedial occipitotemporal cortex (Figure 6) (Bruffaerts et al., 2013a; Liuzzi et al., 2015). These pathways provide afferents to the triangular part which is a convergence zone. There may be intermediary or parallel steps from left perirhinal cortex to the left triangular part that remained below the detection threshold of our method. A candidate region would be the anterior temporal cortex anterior and lateral from the perirhinal cortex.

Potential limitations

The pars triangularis was not present in the logical conjunction analysis, despite the fact that it was the only region that showed a crossmodal semantic similarity effect. As the time activity curves convincingly show (Supplementary Figure 2), its absence from the conjunction analysis was purely a threshold effect. The time activity curve obtained from the univariate analysis by no means contradicts the findings obtained by means of the RSA as it shows activation during the property verification task both for written and for auditory words compared to the control task. The current data illustrate the limitations of relying purely on conjunction analyses, both with regards to type I and type II errors. More importantly, a region can be significant during a semantic task in a conjunction analysis between input modalities while its activity pattern may not reflect the content of the entity, as was the case for several regions examined.

Some regions belonging to the distributed semantic network were not significantly activated during property verification in the VOI defining experiment, for instance, the left anterior temporal pole (Vandenberghe et al., 1996; Lambon Ralph et al., 2010), the angular gyrus (Bonner et al., 2013; Price et al., 2015), and the posterior middle temporal gyrus (Vandenberghe et al., 1996; Whitney et al., 2012; Fairhall and Caramazza, 2013) (for review see Binder et al. (2009)).

The left angular gyrus was more active during the control task than during the property verification task. The level of processing required during the control task was much lower than that required during the property verification task and higher activity in the default mode network, including the left angular gyrus, is a common and highly consistent finding under these conditions (Mason et al., 2007; Lambon Ralph et al., 2016). Given that it has been implicated in semantic processing in previous studies (Binder et al., 2009; Seghier, 2013), the effect of the semantic similarity matrix on the left angular gyrus activity pattern was also examined. No such effect was present. Obviously this does not exclude that the angular gyrus contributes to semantic processing. Semantic functions that have been assigned to the angular gyrus are combinatorial-semantic procssing (Bonner and Price, 2013; Price et al., 2015) and the processing of event concepts rather than single concrete entities (Binder and Desai, 2011). Under these hypotheses one would not necessarily expect an effect of semantic similarity between entities upon its activity pattern.

Left posterior middle temporal gyrus is a hub (Vandenberghe et al., 2013; Xu et al., 2016) within the semantic memory network but its involvement may depend on semantic control demands. For instance, in the absence of a task, the left middle temporal gyrus is not activated during conscious perception of words versus subliminal processing (Van Doren et al., 2010b). The posterior middle temporal gyrus was not activated during the property verification task in the current experiment and neither was it activated during the control condition versus baseline (Figure 4E). The absence of posterior middle temporal gyrus may be due to the relatively low cognitive control demands of our task (Van Doren et al., 2010a; Whitney et al., 2011, 2012).

Neither is the absence of middle temporal gyrus activation anterior and lateral of perirhinal cortex due to activation during the control condition. While the anterior superior temporal gyrus and sulcus were strongly activated during the property verification with auditory words compared to the control condition, there was no activation in the anterior middle and inferior temporal gyrus during the property verification compared to control nor during the control condition compared to baseline. The anterior temporal pole is most consistently activated during fMRI tasks with high combinatorial demands, such as associative-semantic judgments (Vandenberghe et al., 1996) or sentence processing (Vandenberghe et al., 2002). The absence of activation could be due to the low combinatorial demands of the property verification task. Furthermore, in the anterior middle and inferior temporal gyrus TSNR was also lower (Figure 3) than in superior temporal gyrus.

The precuneus and the pars orbitalis were significantly activated in the conjunction analysis of the property verification versus the control task for written and for auditory modality, yet did not exhibit a cross-modal semantic similarity effect. Obviously, conjoint activation does not imply a representation of the meaning across modalities. For instance, semantic control processes may also be conjointly activated for the two input-modalities. The absence of a cross-modal semantic similarity effect in precuneus may appear at odds with a previous MVPA by Fairhall and Caramazza (2013). The latter study of written words and pictures was restricted to the visual input modality, which may explain the difference in outcome with our study. For pairs of written words, the semantic similarity effect showed a trend (Supplementary Table 3).

The absence of semantic similarity effects in regions such as the left angular gyrus, the precuneus, and the left posterior middle temporal gyrus in the current study should by no means be considered as evidence that there cannot be any effect of semantic similarity in these regions under any condition. It has been shown before that the similarity of activity patterns within these regions may correlate with semantic similarities between entities. The main finding of the current paper is the presence of a strong and consistent crossmodal semantic similarity effect in the pars triangularis but this by no means excludes that under other experimental conditions, with different steps of selecting clusters for RSA, or different multivariate encoding or decoding methods (Naselaris et al., 2011) other regions may also exhibit amodal semantic effects (Devereux et al., 2013; Fairhall and Caramazza, 2013; Handjaras et al., 2016).

Conclusion

In summary, our findings show for the first time the effect of semantic similarity on activity patterns across input modalities in the left pars triangularis.

5 Acknowledgements

Liuzzi, Antonietta Gabriella, et al. "Cross-modal representation of spoken and written word meaning in left pars triangularis." NeuroImage 150 (2017): 292-307.

References

Allport DA, Funnell E (1981) Components of the Mental Lexicon. *Philosophical Transactions of the Royal Society B: Biological Sciences* 295:397–410.

Amunts K, Lenzen M, Friederici AD, Schleicher A, Morosan P, Palomero-Gallagher N, Zilles K (2010) Broca's region: novel organizational principles and multiple receptor mapping. *PLoS Biol* 8:e1000489.

Amunts K, Schleicher A, Bürgel U, Mohlberg H, Uylings H, Zilles K (1999) Broca's region revisited: cytoarchitecture and intersubject variability. *Journal of Comparative Neurology* 412:319–341.

Amunts K, Weiss PH, Mohlberg H, Pieperhoff P, Eickhoff S, Gurd JM, Marshall JC, Shah NJ, Fink GR, Zilles K (2004) Analysis of neural mechanisms underlying verbal fluency in cytoarchitectonically defined stereotaxic spacethe roles of Brodmann areas 44 and 45. *Neuroimage* 22:42–56.

Baars BJ (1993) *A cognitive theory of consciousness* Cambridge University Press.

Baars BJ (2002) The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences* 6:47–52.

Baayen RH, Piepenbrock R, Gulikers L (1995) The CELEX lexical database (release 2)[cd-rom]. Philadelphia, PA: Linguistic data consortium, University of Pennsylvania. *Approaches to Human Cognition* pp. 189–225.

Baayen R, Piepenbrock R, Van Rijn H (1993) The CELEX lexical database, 1993. *Philadelphia: University of Pennsylvania* .

Barsalou LW (2016) On staying grounded and avoiding quixotic dead ends. *Psychonomic bulletin & review* pp. 1–21.

Binder JR, Desai RH (2011) The neurobiology of semantic memory. *Trends in cognitive sciences* 15:527–36.

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. *Cerebral cortex* 19:2767–96.

Bonner MF, Peelle JE, Cook PA, Grossman M (2013) Heteromodal conceptual processing in the angular gyrus. *NeuroImage* 71:175–86.

Bonner MF, Price AR (2013) Where Is the Anterior Temporal Lobe and What Does It Do? *Journal of Neuroscience* 33:4213–4215.

Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM (2002) Modality independence of word comprehension. *Human brain mapping* 16:251–61.

Bruffaerts R, Dupont P, De Grauwe S, Peeters R, De Deyne S, Storms G, Vandenberghe R (2013a) Right fusiform response patterns reflect visual object identity rather than semantic similarity. *NeuroImage* 83:87–97.

Bruffaerts R, Dupont P, Peeters R, De Deyne S, Storms G, Vandenberghe R (2013b) Similarity of fMRI activity patterns in left perirhinal cortex reflects semantic similarity between words. *The Journal of neuroscience* 33:18597–607.

Bundesen C, Habekost T, Kyllingsbaek S (2005) A neural theory of visual attention: bridging cognition and neurophysiology. *Psychological review* 112:291–328.

Caramazza A, Hillis A, Rapp B, Romani C (1990) The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology* 7:161–189.

Chafe W, Tannen D (1987) The Relation Between Written and Spoken Language. *Annual Review of Anthropology* 16:383–407.

Chee MW, Caplan D, Soon CS, Sriram N, Tan EW, Thiel T, Weekes B (1999) Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron* 23:127–37.

Constable R, Pugh KR, Berroya E, Mencl W, Westerveld M, Ni W, Shankweiler D (2004) Sentence complexity and input modality effects in sentence comprehension: an fMRI study. *NeuroImage* 22:11–21.

De Deyne S, Navarro DJ, Perfors A, Storms G (2016) Structure at every scale: A semantic network account of the similarities between unrelated concepts. *Journal of Experimental Psychology: General* 145:1228.

De Deyne S, Navarro DJ, Storms G (2013) Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods* 45:480–498.

De Deyne S, Verheyen S, Ameel E, Vanpaemel W, Dry MJ, Voorspoels W, Storms G (2008) Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior research methods* 40:1030–48.

De Deyne S, Verheyen S, Storms G (2015) The role of corpus size and syntax in deriving lexico-semantic representations for a wide range of concepts. *The Quarterly Journal of Experimental Psychology* 68:1643–1664.

Demb J, Desmond J, Wagner A, Vaidya C, Glover G, Gabrieli J (1995) Semantic encoding and retrieval in the left inferior prefrontal cortex: a functional MRI study of task difficulty and process specificity. *J. Neurosci.* 15:5870–5878.

Devereux BJ, Clarke A, Marouchos A, Tyler LK (2013) Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *The Journal of neuroscience* 33:18906–16.

Devlin JT, Matthews PM, Rushworth MF (2003) Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *Journal of cognitive neuroscience* 15:71–84.

Duncan J, Owen A (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* 23:475–483.

Duncan J (2001) An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience* 2:820–829.

Eickhoff SB, Heim S, Zilles K, Amunts K (2006) Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *Neuroimage* 32:570–582.

Eickhoff SB, Paus T, Caspers S, Grosbras MH, Evans AC, Zilles K, Amunts K (2007) Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *Neuroimage* 36:511–521.

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.

Fairhall SL, Caramazza A (2013) Brain Regions That Represent Amodal Conceptual Knowledge. *Journal of Neuroscience* 33:10552–10558.

Fernandino L, Binder JR, Desai RH, Pendl SL, Humphries CJ, Gross WL, Conant LL, Seidenberg MS (2015) Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cerebral Cortex* p. bhv020.

Frey S, Campbell JSW, Pike GB, Petrides M (2008) Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *The Journal of neuroscience* 28:11435–44.

Friederici AD (2002) Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6:78–84.

Friederici AD (2011) The brain basis of language processing: from structure to function. *Physiological reviews* 91:1357–92.

Friederici AD, Gierhan SME (2013) The language network. *Current opinion in neurobiology* 23:250–4.

Friston KJ, Holmes aP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ (1995) Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping* 2:189–210.

Gabrieli JD, Poldrack RA, Desmond JE (1998) The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences of the United States of America* 95:906–913.

Gerbella M, Belmalih A, Borra E, Rozzi S, Luppino G (2010) Cortical connections of the macaque caudal ventrolateral prefrontal areas 45A and 45B. *Cerebral cortex* 20:141–68.

Gold BT, Balota DA, Cortese MJ, Sergent-Marshall SD, Snyder AZ, Salat DH, Fischl B, Dale AM, Morris JC, Buckner RL (2005a) Differing neuropsychological and neuroanatomical correlates of abnormal reading in early-stage semantic dementia and dementia of the Alzheimer type. *Neuropsychologia* 43:833–46.

Gold BT, Balota DA, Kirchhoff BA, Buckner RL (2005b) Common and dissociable activation patterns associated with controlled semantic and phonological processing: evidence from fMRI adaptation. *Cerebral cortex* 15:1438–50.

Goldman-Rakic PS (1987) Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. *Comprehensive Physiology* .

Goucha T, Friederici AD (2015) The language skeleton after dissecting meaning: A functional segregation within Broca's Area. *NeuroImage* 114:294–302.

Hagoort P (2005) On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences* 9:416–423.

Hamzei F, Vry MS, Saur D, Glauche V, Hoeren M, Mader I, Weiller C, Rijntjes M (2016) The dual-loop model and the human mirror neuron system: an exploratory combined fMRI and dti study of the inferior frontal gyrus. *Cerebral Cortex* 26:2215–2224.

Handjaras G, Ricciardi E, Leo A, Lenci A, Cecchetti L, Cosottini M, Marotta G, Pietrini P (2016) How concepts are encoded in the human brain: A modality independent, category-based cortical organization of semantic knowledge. *NeuroImage* 135:232–242.

Heim S, Alter K, Ischebeck AK, Amunts K, Eickhoff SB, Mohlberg H, Zilles K, von Cramon DY, Friederici AD (2005) The role of the left Brodmann's areas 44 and 45 in reading words and pseudowords. *Cognitive Brain Research* 25:982–993.

Hickok G, Poeppel D (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92:67–99.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nature Reviews Neuroscience* 8:393–402.

Homae F (2002) From perception to sentence comprehension: The Convergence of auditory and visual information of language in the left inferior frontal cortex. *NeuroImage* 16:883–900.

Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532:453–458.

Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76:1210–24.

Katzev M, Tüscher O, Hennig J, Weiller C, Kaller CP (2013) Revisiting the functional specialization of left inferior frontal gyrus in phonological and semantic fluency: the crucial role of task demands and individual ability. *The Journal of neuroscience* 33:7837–45.

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. *Human brain mapping* 30:4116–28.

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in systems neuroscience* 2:4.

Lambon Ralph MA, Cipolotti L, Manes F, Patterson K (2010) Taking both sides: Do unilateral anterior temporal lobe lesions disrupt semantic memory? *Brain* 133:3243–3255.

Lambon Ralph MA, Jefferies E, Patterson K, Rogers TT (2016) The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience* .

Lau EF, Phillips C, Poeppel D (2008) A cortical network for semantics:(de) constructing the n400. *Nature Reviews Neuroscience* 9:920–933.

Levenshtein V (1966) Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10:707 – 710.

Liuzzi AG, Bruffaerts R, Dupont P, Adamczuk K, Peeters R, De Deyne S, Storms G, Vandenberghe R (2015) Left perirhinal cortex codes for similarity in meaning between written words: Comparison with auditory word input. *Neuropsychologia* pp. 1–13.

Mason MF, Norton MI, Van Horn JD, Wegner DM, Grafton ST, Macrae CN (2007) Wandering minds: the default network and stimulus-independent thought. *Science* 315:393–395.

McDermott KB, Petersen SE, Watson JM, Ojemann JG (2003) A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia* 41:293–303.

Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. *Neuroimage* 56:400–410.

Petrides M, Pandya DN (2002) Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *European Journal of Neuroscience* 16:291–310.

Petrides M, Pandya DN (2009) Distinct Parietal and Temporal Pathways to the Homologues of Broca's Area in the Monkey. *PLoS Biology* 7:e1000170.

Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD (1999) Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10:15–35.

Poline JB, Worsley KJ, Evans AC, Friston KJ (1997) Combining spatial extent and peak intensity to test for activations in functional imaging. *NeuroImage* 5:83–96.

Price AR, Bonner MF, Peelle JE, Grossman M (2015) Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *The Journal of neuroscience* 35:3276–84.

Price CJ, Wise RJ, Frackowiak RS (1996) Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral cortex* 6:62–70.

Rijntjes M, Weiller C, Bormann T, Musso M (2012) The dual loop model: its relation to language and other modalities. *Frontiers in evolutionary neuroscience* 4:9.

Saleem KS, Kondo H, Price JL (2008) Complementary circuits connecting the orbital and medial prefrontal networks with the temporal, insular, and opercular cortex in the macaque monkey. *The Journal of comparative neurology* 506:659–93.

Sass K, Krach S, Sachs O, Kircher T (2009) Lion-tiger-stripes: neural correlates of indirect semantic priming across processing modalities. *Neuroimage* 45:224–236.

Saur D, Kreher BW, Schnell S, Kümmerer D, Kellmeyer P, Vry MS, Umarova R, Musso M, Glauche V, Abel S, Huber W, Rijntjes M, Hennig J, Weiller C (2008) Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America* 105:18035–40.

Saur D, Schelter B, Schnell S, Kratochvil D, Küpper H, Kellmeyer P, Kümmerer D, Klöppel S, Glauche V, Lange R, Mader W, Feess D, Timmer J, Weiller C (2010) Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. *NeuroImage* 49:3187–97.

Scott SK (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123:2400–2406.

Seghier ML (2013) The angular gyrus: multiple functions and multiple subdivisions. *The Neuroscientist* 19:43–61.

Simanova I, Hagoort P, Oostenveld R, van Gerven MAJ (2014) Modality-independent decoding of semantic information from the human brain. *Cerebral cortex* 24:426–34.

Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJS (2006) Converging language streams in the human temporal lobe. *The Journal of neuroscience* 26:7328–36.

Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2013) Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78:364–375.

Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ (1997) Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences of the United States of America* 94:14792–14797.

Ueno T, Saito S, Rogers TT, Lambon Ralph MA (2011) Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron* 72:385–96.

Uylings H, Malofeeva L, Bogolepova I, Amunts K, Zilles K (1999) Brocas language area from a neuroanatomical and developmental perspective. *The neurocognition of language* pp. 319–336.

Van Doren L, Dupont P, De Grauwe S, Peeters R, Vandenberghe R (2010a) The amodal system for conscious word and picture identification in the absence of a semantic task. *NeuroImage* 49:3295–307.

Van Doren L, Dupont P, De Grauwe S, Peeters R, Vandenberghe R (2010b) 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* 383:254–6.

Vandenberghe R, Nobre AC, Price C (2002) The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience* 14:550–560.

Vandenberghe R, Wang Y, Nelissen N, Vandenbulcke M, Dhollander T, Sunaert S, Dupont P (2013) The associative-semantic network for words and pictures: Effective connectivity and graph analysis. *Brain and Language* 127:264–272.

Vandenbulcke M, Peeters R, Fannes K, Vandenberghe R (2006) Knowledge of visual attributes in the right hemisphere. *Nat. Neurosci.* 9:964–970.

Wagner AD (2000) Task-specific Repetition Priming in Left Inferior Prefrontal Cortex. *Cerebral Cortex* 10:1176–1184.

Wagner AD, Paré-Blagoev E, Clark J, Poldrack RA (2001) Recovering Meaning. *Neuron* 31:329–338.

Weiller C, Bormann T, Saur D, Musso M, Rijntjes M (2011) How the ventral pathway got lost–and what its recovery might mean. *Brain and language* 118:29–39.

Whitney C, Kirk M, O'Sullivan J, Lambon Ralph MA, Jefferies E (2011) The Neural Organization of Semantic Control: TMS Evidence for a Distributed Network in Left Inferior Frontal and Posterior Middle Temporal Gyrus. *Cerebral Cortex* 21:1066–1075.

Whitney C, Kirk M, O'Sullivan J, Lambon Ralph MA, Jefferies E (2012) Executive semantic processing is underpinned by a large-scale neural network: revealing the contribution of left prefrontal, posterior temporal, and parietal cortex to controlled retrieval and selection using TMS. *Journal of cognitive neuroscience* 24:133–47.

Xu Y, Lin Q, Han Z, He Y, Bi Y (2016) Intrinsic functional network architecture of human semantic processing: Modules and hubs. *NeuroImage* 132:542–555.