

## Journal Pre-proof

Distinct fronto-temporal substrates of distributional and taxonomic similarity among words: evidence from RSA of BOLD signals

Francesca Carota Interpretation , Hamed Nili ,  
Friedemann Pulvermüller , Nikolaus Kriegeskorte

PII: S1053-8119(20)30893-4  
DOI: <https://doi.org/10.1016/j.neuroimage.2020.117408>  
Reference: YNIMG 117408



To appear in: *NeuroImage*

Received date: 16 May 2020  
Revised date: 25 September 2020  
Accepted date: 28 September 2020

Please cite this article as: Francesca Carota Interpretation , Hamed Nili , Friedemann Pulvermüller , Nikolaus Kriegeskorte , Distinct fronto-temporal substrates of distributional and taxonomic similarity among words: evidence from RSA of BOLD signals, *NeuroImage* (2020), doi: <https://doi.org/10.1016/j.neuroimage.2020.117408>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Inc.  
This is an open access article under the CC BY-NC-ND license  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## **Distinct fronto-temporal substrates of distributional and taxonomic similarity among words: evidence from RSA of BOLD signals**

Francesca Carota<sup>1,2,3,4,5</sup>, Hamed Nili<sup>3,6</sup>, Friedemann Pulvermüller<sup>3,4,5</sup> & Nikolaus Kriegeskorte<sup>3,7</sup>

<sup>1</sup> Max-Planck-Institute for Psycholinguistics, Wundtlaan 1, Nijmegen, the Netherlands

<sup>2</sup> Donders Institute for Cognitive NeuroImaging, Kapittelweg 29, 6525 EN Nijmegen, the Netherlands

<sup>3</sup> MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 7EF United Kingdom

<sup>4</sup> Berlin School of Mind and Brain, Humboldt Universität zu Berlin, Berlin, Germany

<sup>5</sup> Brain Language Laboratory, Department of Philosophy and Humanities, WE4, Freie Universität Berlin, Berlin, Germany

<sup>6</sup> Department of Experimental Psychology, University of Oxford, Tinbergen Building, 9 South Parks Road, Oxford OX1 3UD, UK

<sup>7</sup> Cognitive Imaging at the Zuckerman Mind Brain Behavior Institute, Columbia University, Jerome L. Greene Science Center, 3227 Broadway, L3-064, 9834 New York, NY 10027

Corresponding author: Francesca Carota, Berlin School of Mind and Brain, Humboldt Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany *Email to: [francesca.carota@mpi.nl](mailto:francesca.carota@mpi.nl) or [f.carota@donders.ru.nl](mailto:f.carota@donders.ru.nl)*

### **Abstract**

A class of semantic theories defines concepts in terms of statistical distributions of lexical items, basing meaning on vectors of word co-occurrence frequencies. A different approach emphasizes abstract hierarchical taxonomic relationships among concepts. However, the functional relevance of these different accounts and how they capture information-encoding of meaning in the brain still remains elusive.

We investigated to what extent distributional and taxonomic models explained word-elicited neural responses using cross-validated representational similarity analysis (RSA) of functional magnetic resonance imaging (fMRI) and novel model comparisons.

Our findings show that the brain encodes both types of semantic similarities, but in distinct cortical regions. Posterior middle temporal regions reflected word links based on hierarchical taxonomies, along with the action-relatedness of the semantic word categories. In contrast, distributional semantics best predicted the representational patterns in left inferior frontal gyrus (LIFG, BA 47). Both representations coexisted in angular gyrus supporting semantic binding and integration. These results reveal that neuronal networks with distinct cortical distributions across higher-order association cortex encode different representational properties of word meanings. Taxonomy may shape long-term lexical-semantic representations in memory consistently with sensorimotor details of semantic categories, whilst distributional knowledge in the LIFG (BA 47) enable semantic combinatorics in the context of language use.

Our approach helps to elucidate the nature of semantic representations essential for understanding human language.

**Keywords:** conceptual taxonomies, co-occurrence statistics, fMRI, representational similarity searchlights, language comprehension.

## **Introduction**

Human language arises from the combination of meaningful building blocks: words. A range of cortical regions becomes active when human subjects process word meanings and, therefore, semantic comprehension has been linked to widespread brain systems (Pulvermüller, 1999; Binder et al., 2009). Still, there is disagreement about the specific roles of these regions in representing lexical meaning. "Hub-and-spokes" models posit a single semantic "hub" region in anterior temporal lobe (aTL; Patterson et al., 2007; Lambon-Ralph, 2017), to integrate motor, auditory and visual semantic features processed in modality-specific cortex into coherent multimodal semantic representations of concepts. Different accounts view several regions in

multimodal association cortex (henceforth multimodal regions) as bases of integrated representations derived from multiple low-level sensory and motor representations (Damasio, 1989a, 1989b; Damasio et al., 1996; Price et al., 2004; Mesulam et al., 1995; Martin et al., 2007; Meyer and Damasio, 2009; Binder and Desai 2011; Pulvermüller, 2013). Furthermore, a role of modality-preferential visual and motor cortex has been highlighted in the context of semantic grounding for specific categories of object and action words (Barsalou, 2008; Kiefer and Pulvermüller, 2012; Carota et al., 2012, 2017). For example, recent TMS results reinforced the view of a causal contribution of the motor systems in lexical-semantic encoding, bringing evidence for category-specific motor processes also when action words are learnt (Vukovic and Shtyrov, 2019).

To determine the regions of semantic representation, neuroimaging experiments measuring brain spatial-averaged activations to meaningful stimuli are insufficient, because they reveal “where” words make sense in the brain, but not the types of computations performed in specific areas, i.e. how the brain represents information to achieve the semantic interpretation of words and concepts. The brain represents concepts by patterns of activity across populations of neurons. Multivariate pattern-information analyses offer a powerful tool to reveal the information these population codes carry by measuring the variation of activity across multiple neighbouring neuronal units (e.g. neurons or voxels) (e.g., Haxby et al., 2001; Hanson et al., 2004; Kriegeskorte, 2004; Mitchell et al., 2004; Polyn et al., 2005; Kiani et al., 2007; Kriegeskorte and Bandettini 2007 for review). Recent multivariate imaging studies have demonstrated that the brain encodes the semantic similarity between concepts in terms of similarity (e.g. correlation distance) between these activation patterns (e.g. Kriegeskorte et al.,

2008; Devereux et al., 2013; Carlson et al., 2014; Carota et al., 2017; Mitchell and Cusack, 2016), thus opening new perspectives for mapping brain semantic content.

In this context, a critical factor influencing multivariate semantic brain mapping results could be the way conceptual similarity is measured. Distributional theories define the meaning of a word on the basis of the other words it frequently co-occurs with in the actual context of language use (Harris, 1954; Firth, 1957; Miller, 1967). However, an entirely different approach, which is immanent to some psycholinguistic models (Shiffrin, 1970; Miller, 1995; Fellbaum, 1998) and is elaborated in the WordNet lexical database, exploits abstract hierarchical-taxonomic relations to capture the semantic similarity between two words (e.g. *snake* and *moose*) based on their common feature of representing animals (e.g. both animals are vertebrates, but one is a reptile whereas the other is a mammal). As these models use different criteria, it is not surprising that related imaging results about the brain substrates of semantic similarity differ.

Importantly, these different accounts tap into key organisational principles of semantic knowledge, structuring the relationships among the concepts which words refer to. In particular, hierarchical structure reflects word referential meaning, the "word-world" links between word forms and the objects, actions and entities in the extralinguistic world (de Saussure, 1916). In contrast, distributional statistics capture "word-word" relations as induced by language contexts (e.g., texts, media, discourse), which are an essential source of semantic learning and understanding, since words which have not been learnt in the actual presence of an object or action (e.g.: "peach" in the physical context of a peach) can still be understood by inferring their meaning from the already known words they co-occur with (Landauer and Dumais, 1998; also see Carota et al., 2017). Although distributional approaches capture an important aspect of semantic information, it is still possible that they do not reflect the neural representation of

conceptual knowledge (Glenberg and Mehta, 2008) and that distributional/contextual and hierarchical models focus on different semantic computations (Maki and Buchanan, 2008). Therefore, comparing these two types of models can be particularly insightful for investigating cortical information processing of words.

Earlier studies using multi-voxel pattern analyses (MVPA) reported evidence of semantic similarity among words and concepts based on hierarchical/taxonomic relations in regions of higher-level visual cortex (e.g., Fairhall and Caramazza, 2013; Mitchell and Cusack, 2016). In contrast, other previous MVPA work suggested that distributional models may explain the representational patterns in higher-level visual regions even better than the taxonomy (Carlson et al., 2014). Recent neuroimaging results have also suggested distinct brain correlates of hierarchical and distributional relations in the aTL and the inferior parietal cortex, respectively (Xu et al., 2016). A different set of results has shown that distributional links could be mapped onto distributed regions, including visual and motor cortex (Mitchell et al., 2008), and left inferior frontal cortex (e.g., Carota et al., 2017), thus encompassing both multimodal fronto-temporal cortex and modality-preferential motor regions (e.g., Pereira et al., 2018). Therefore, these previous advances in identifying the neural bases of semantic comprehension did not clarify how these qualitatively different semantic dimensions relate to word representations in the semantic systems.

In the present study, we asked whether one approach is better suited than the other for interpreting the semantic information carried by brain response patterns. If both map brain indices of semantic similarity, a second question is whether partly distinct, possibly complementary brain regions contribute to the semantic aspects captured by these models (Carota et al., 2016; 2017). To address these questions, we investigated and systematically

compared the brain representations of individual word meanings captured by these different semantic models throughout the cortex.

Since distributional metrics capture the statistical probability of words to co-occur in language usage regardless their semantic word category, we hypothesized that distributional semantics would relate with activity in neuronal assemblies distributed across multimodal semantic regions (Pulvermüller, 2013), encompassing more anterior sections of the left inferior frontal gyrus (LIFG BA 45-47) (Carota et al., 2017; Garagnani and Pulvermüller, 2016; Pulvermüller, 2013), the aTL (Patterson et al., 2007; Lambon-Ralph, 2017), middle temporal gyrus (MTG, Turken and Dronkers, 2011) and inferior parietal cortex (especially angular gyrus, AG, BA 39: Binder et al., 2009).

In contrast, we expected that the taxonomic relationships among words from similar semantic categories yielded correlations with the similarity patterns in category-preferential regions in posterior middle and inferior temporal gyrus (pMTG, pITG) and motor cortex, in which the sensory and motor features of the corresponding referential meanings are grounded (Barsalou, 2008; Pulvermüller, 2013). Consequently, we expected that, in these same posterior temporal and motor regions, the neural patterns reflecting hierarchical/taxonomic relations would also conform to the action-specific motor properties of our set of action-related words, as assessed by behavioural ratings.

To test these hypotheses, we presented 96 words in a functional magnetic resonance imaging (fMRI) experiment and used representational similarity analysis (RSA) searchlights and novel model comparisons (Kriegeskorte et al., 2008; Nili et al., 2014), to relate the word-elicited multivoxel fMRI patterns to distributional semantic, as quantified by a state-of-the-art vector-

embedding language model (Word2Vec, Mikolov, et al., 2013), and hierarchical semantic models as captured by WordNet.

## Materials and Methods

**Participants.** Twenty-three healthy volunteers participated in the study. All participants were right-handed (laterality quotient of 90, standard error (SE) = 3.1), monolingual English native speakers (mean age 29 years, SE=2.8). Participants had no history of neurological or psychiatric disorders. They had normal or corrected-to-normal vision. All participants gave their informed consent to take part in the study and were remunerated for their time. Ethical approval was obtained from the Cambridge Psychology Research Ethics Committee.

**Stimuli.** Ninety-six words, sixteen from each individual category of leg-, arm-, face-related actions and tool-, animal-, food-related objects, were selected based on established semantic ratings (Pulvermüller et al., 1999, Carota et al., 2012). Stimulus word groups were matched for a range of psycholinguistic properties, including word length (counted in number of letters), letter bigram and trigram frequency, logarithmic word frequency, number of orthographic neighbors, and standardized lexical frequency, while differing in imageability, concreteness, and action-relatedness (see Table 1). We obtained relevant values from the CELEX database (Baayen et al. 1993) and the WordSmyth Website ([www.wordsmyth.net/](http://www.wordsmyth.net/)). 21% of the action words were lexically unambiguous verbs and the lexically ambiguous ones which could be used as nouns and verbs were in the average 14 times more frequently used as verbs than as nouns (according to the CELEX database: Baayen et al. 1993; SE 4.2). 58% of the object words were lexically



unambiguous nouns and the lexically ambiguous ones which could be used as nouns and verbs were in the average 6 times more frequently used as nouns than as verbs (SE 2). Strings of meaningless hash marks matched in length to the stimulus words were used as low-level baseline stimuli during 120 trials. Null events were presented during 60 trials. Sixty trials consisting of misspelled words to be detected by the participants throughout the experimental task were presented. These “typo” trials (e.g. feele instead of feel, heate instead of heat, branc instead of branch, lurch instead of lunch) did not include words from any of the semantic categories from which the 96 target words were taken - so as to avoid a bias towards one of these categories -and were discarded from the analysis. After the fMRI experiment, participants completed an unannounced word recognition test containing both novel distractor and experimental words, to further confirm that they had attended to the word reading task. They performed above chance (average hit rate: 80% [STD: 8.3%]), indicating their attention to the words and compliance with the task.

**Experimental Design.** We adopted a rapid, periodic single trial, event-related paradigm. Stimulus duration was 100 msec. A fixation cross was presented at the centre of the screen between two consecutive stimuli. The stimulus onset asynchrony (SOA) was jittered ~3.5-4 s. The 96 stimulus words were presented in a different pseudo-random order in each of the 6 runs. Each stimulus occurred once per run. Stimuli were visually presented by E-Prime software (Psychology Software Tools, Inc., Sharpsburg, PA) through a back-projection screen positioned in front of the scanner and viewed on a mirror placed on the head coil.

**Task.** Participants were engaged in an attentive reading task with occasional typo-detection

(~7% of trials). They received the instruction to attend to all stimuli, to silently read the words and to understand their meanings. In addition, we instructed the participants to press a button with their left hand if a misspelled word appeared at the centre of the screen. We chose attentive reading with no semantic task in order to avoid semantic processing strategies confounding the processes this study aims at, namely the semantic understanding of the target words (see Discussion). We included in the task the occasional typo-detection to maintain subject alertness and we specifically chose it for being orthogonal to the main task and not interfering with semantic comprehension.

**Imaging Methods.** Participants were scanned in a Siemens 3T Tim Trio using a head coil (12 channels to receive). Echo-planar imaging (EPI) sequence parameters were TR = 2000 msec, TE = 30msec, and flip angle = 78 degrees. The functional images consisted of 32 slices covering the whole brain (slice thickness 3mm, in-plane resolution 3mm x 3mm, inter-slice distance 0.75mm).

**Data Analysis.** Imaging data were analysed using SPM12 software (Wellcome Department of Imaging Neuroscience, London, UK). Images were corrected for slice timing and re-aligned to the first image using sinc interpolation. The EPI images were co-registered to the structural T1 images using a mutual coregistration procedure (Maes et al., 1997). The structural MRI was normalised to the 152-subject T1 template of the Montreal Neurological Institute (MNI). The resulting transformation parameters were applied to the co-registered EPI images.

**Representational Similarity Analysis.** For multivariate RSA (Kriegeskorte et al., 2008; Nili et al., 2014), the analysis was carried out in participant native space, using realigned, unsmoothed

and non-normalised functional data, which were co-registered with MPRAGE of each subject. Data were analysed using the general linear model. Response-amplitude was estimated for each voxel and for each of the 96 stimuli by performing single univariate linear model fit. Runs were concatenated along the temporal dimension. A separate hemodynamic predictor was included for each of the 96 stimulus words. The time course of the predictors was determined based on the event sequence and a linear model of the hemodynamic response (Boynton et al., 1996). For each run, the design matrix was composed of the stimulus-response predictors with six head motion parameter time courses and a confound-mean predictor. The response-amplitude (beta) estimate map associated with each stimulus was converted into a t map by contrasting them against the implicit baseline in order to compute the RDMs (Kriegeskorte et al., 2008; Nili et al., 2014).

### *fMRI data analyses*

**Whole-brain Searchlights RSA.** Data were extracted for each participant individually using a “sphere of information” searchlight approach (Kriegeskorte et al., 2008; Nili et al., 2014). A roaming spherical searchlight with 10 mm radius was moved throughout the grey matter to extract continuous, voxel-by-voxel maps of word-elicited activation values. To achieve maximal sensitivity to our experimental manipulations, this analysis was based on single items, with each experimental word modeled as a condition and associated with a separate hemodynamic predictor. The correlation distances (1-Pearson’s correlation) between the response patterns for each word paired with every other word were expressed as representational dissimilarity matrices (RDMs), which are symmetric about a diagonal of zeros (Kriegeskorte et al., 2008). These brain data RDMs were then correlated with theoretical model RDMs (using Spearman’s rank

correlation) at each brain location. The resulting maps of  $r$  values for each participant and model were normalised onto the MNI template and entered into a group-level random-effects (RFX) analysis using permutation-based non-parametric statistics in SNPM (<http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/software/snpm>), to test for positive correlations between the model RDMs and brain data RDMs and thus determine the brain regions where the models best explained the brain activity. FDR correction at 0.05 for multiple comparisons across voxels and number of models was applied. 10,000 permutations were used in the analysis. In order to ensure that the searchlight maps we reported did not suffer from distortion due to either the searchlight size or the detection of fewer informative voxels, we applied small volume correction testing for activation in spherical ROIs created around the activation peaks for the voxel clusters specific to the taxonomic and distributional model (see Etzel et al. 2013). Three spherical ROIs were centred at coordinates: -54 -61 -1 (pMTG), -39 -66 36 (AG); -33 23 -9 (LIFG, BA 47). Please note that the ROI analyses for model comparisons we report in the main method and results were based on anatomical ROIs, which were independently selected (as described below) to avoid circularity. The method confirmed that the activated voxels we identified with searchlight analyses carried local information related to our effects of interest.

**Data Cross-validation.** To further estimate predictive performance and adjudicate between multiple models, searchlight results were cross-validated using LDC (cross-validated Mahalanobis distance) on split data (Walther et al., 2015). In this approach, for each of  $k$  folds,  $k-1$  of  $k$  independent subsets of the data (training set) are used to fit the parameters of each model and the left-out subset (test-set) is used to estimate predictive performance. Cross-

validated LDC distances provide almost unbiased and conservative estimates of pattern dissimilarity. FDR correction at 0.05 for multiple comparisons across voxels and number of models was applied. 10,000 permutations were used in the analysis.

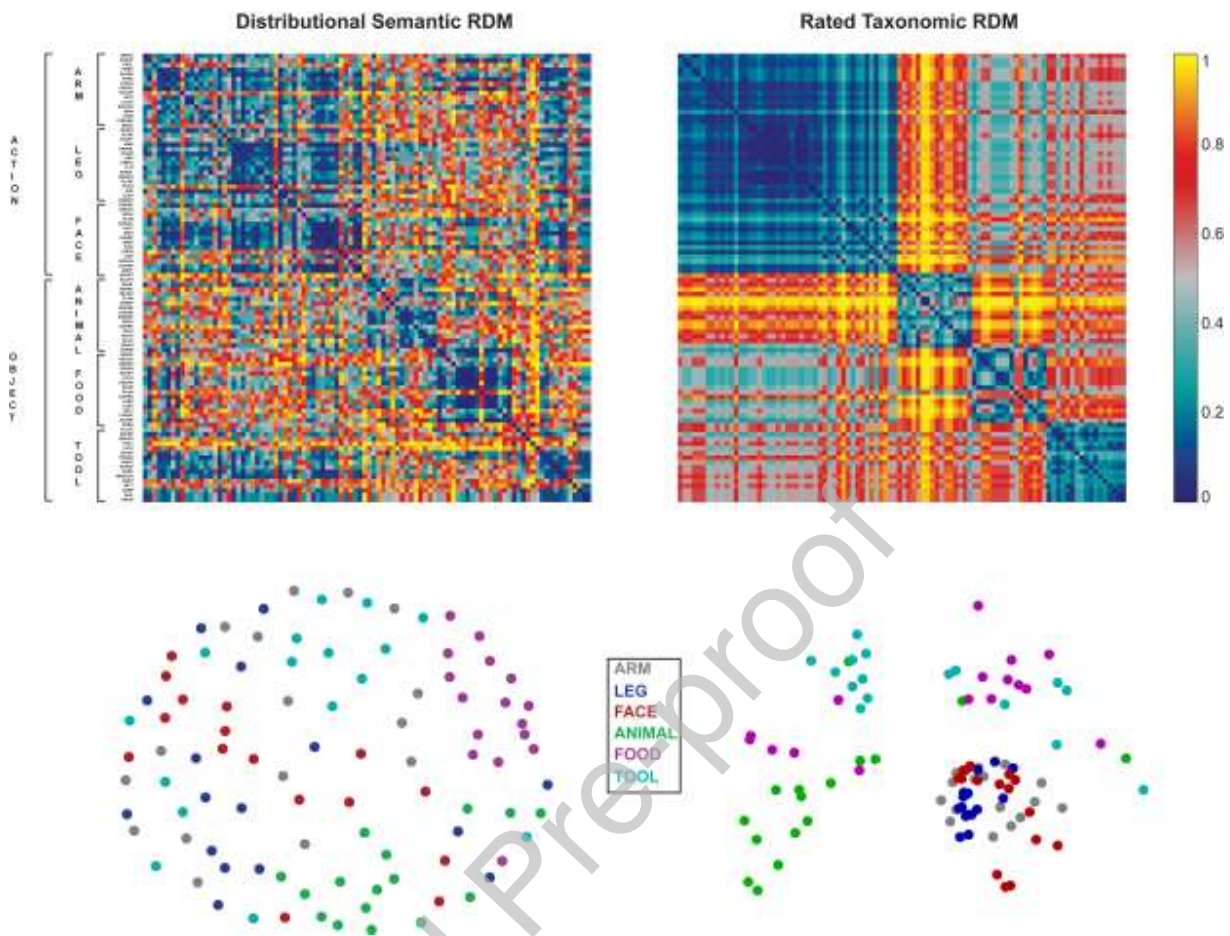
***Statistical model comparisons.*** For each subject, non-parametric Spearman's rank correlations between model and brain activity RDMs were computed in 4 selected ROIs. These included the "semantic" aspect the inferior frontal cortex (Poldrack et al., 1999; Hagoort, 2014), i.e. the pars orbitalis of the LIFG (BA47), the aTL, the pMTG, and the AG. We automatically defined the ROIs using the standard Wake Forest University (WFU) Pickatlas toolbox, which generates ROI masks in standard MNI space based on the Automated Anatomical Labelling (AAL) parcellation. In order to carry out multivariate analysis within individual-subject native space, all ROI-masks were back-projected in each subject's native space by inverting the spatial normalisation applied during GLM analysis. The two models were compared by subtracting the r-value of the correlation between the second model and the fMRI RDM from the r-value of the correlation between the first model and the fMRI RDM. The difference in r-value across all subjects was then tested against the null hypothesis of the value 0, to test for a difference in correlation, using a 2-sided Wilcoxon signed-rank test. P-values surviving FDR correction for multiple comparisons are reported (Benjamini and Hochberg 1995).

### ***Specification of the Linguistic Models***

*Taxonomic model.* A first model aimed at identifying the brain regions representing taxonomic structure, and was based on WordNet (Fellbaum et al., 1998). The organizing criterion of this database is the semantic relation of synonymy. Synonyms - words denoting the same concept - are grouped into unordered sets called synsets, which reach the number of about 117.000 units. Each synset is linked to hierarchies of words overarched by a distinct conceptual node and is linked to other synsets by means of conceptual relations. Here we focused on the "is a" conceptual relation of hypernymy (*Y is a hypernym of X if every X is a kind of Y*), which defines the link between a more general synset like {*furniture*} to more specific ones like {*bed*}. For each two words, the number of edges between the first sense of the target word in the 'is-a' hierarchy was taken to define the distance between two words (see Supplementary Materials for alternative encoding type). The edge count denotes the minimal number of edges needed to traverse from one word to the other in the WordNet hypernym hierarchy, that is: from the first word, via the closest ancestor node common to both words, to the second word. For instance, snake and frog have an edge distance of 5: snake > diapsid > reptile > vertebrate, frog < amphibian < vertebrate.

Because in WordNet each word may contain more than one synset, to avoid bias in the selection criterion, we followed a reviewer's suggestion to invite a group of participants for a rating study motivating the sense choice. Fifteen British English native speakers participated in a rating study, in which the stimulus words were presented in random order. They were asked to read the word, keep the meaning that it evokes in mind, and select that meaning from a randomly ordered list of multiple meanings associated with that word in WordNet. For 75% of the items, there was unanimous agreement (90% of the participants), whilst for the remaining 25% there was no clear dominant meaning assigned. To reflect the distribution of meaning assignments, we therefore

weighted the synsets accordingly, calculating the similarity between two words as the average of the distances between any assigned meaning for one word with any assigned meaning for the other word (see Inline Supplementary Materials). The values of the corresponding taxonomic similarity for each word pair were expressed as RDMs, as shown in Fig. 1 displays the arrangement of the experimental words based on the taxonomic model using multidimensional scaling (MDS). As shown in Fig. 1, one feature of our resulting rated WordNet model was that it expressed cross-category relations among concepts of actions and objects. For example, the model incorporated within-category conceptual relations between *harp* and *drum*, as well as so-called weak (or thematic) associations, like *peach* and *knife*, and *peel* and *knife*. Because of these characteristics, the present model captures both within-category conceptual relatedness and thematic similarity among concepts over and above categorical borders. We consider the implications of this structure for evaluating our results in the discussion.



-----

**Figure 1.** Top panel: The representational dissimilarity matrices (RDMs: 1 = dissimilarity — in yellow -, 0 = similarity, in blue) displaying distributional (on the left) and taxonomic similarities (on the right) among the 96 stimulus words. Bottom panel: arrangement of the experimental stimulus words by multidimensional scaling (MDS), reflecting the distributional semantic distances between all words for the distributional model (on the left) and taxonomic model (on the right). Visual inspection of both RDMs and MDS revealed that the Action words that were related to arm (grey), face (red), and leg (blue) formed a widely interwoven cluster, which also contained the set of object words related to tools (cyan). Object words related to foods (in pink) and animals (in green) formed two separate distributional clusters, which reflected category



boundaries only to some degree. Note that words along the semantic categories of animals, tools, foods, and arm, leg, and face actions did not form well distinct clusters. On the right: MDS reflecting the distributional semantic distances between all words for the taxonomic model. Similar to the RDM visualization of the distributional model, the arrangement by MDS showed that action words, which related to arm (grey), face (red), and leg (blue), formed an interwoven cluster, which was separate from the set of object words. Words related to tools (cyan), foods (in pink) and animals (in green) formed well distinct clusters, suggesting a category-based structure of the similarities expressed by the taxonomic model.

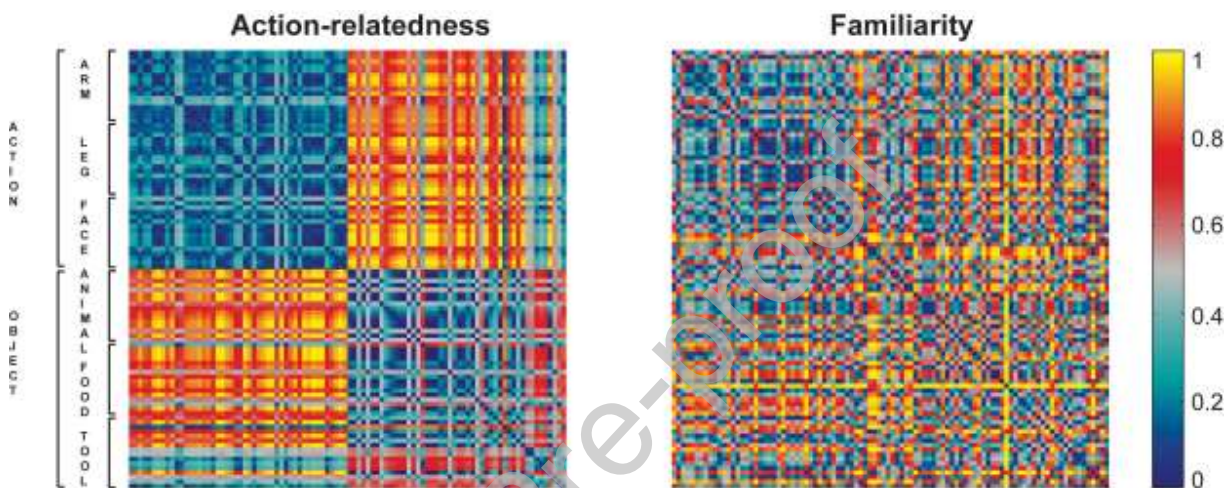
*Distributional model.* To test for the effects of associative semantic relationships on the representation of single word meanings as a function of their statistical distributions, we constructed a computational linguistic model coding for their co-occurrence frequency in texts. Following current state-of-the-art vector-embedding language models, which elaborate classical Latent Semantic Analysis (LSA, Landauer and Dumais, 1998), we constructed such model based on Word2Vec (Mikolov, et al., 2013), based on recent evidence supporting Word2Vec as a reliable model of human semantic representation, allowing to make predictions about behavioural results or human annotations of data (Pereira et al., 2016). Similar to LSA, the model assumes that words have similar meaning if they tend to occur, beyond the *same* textual span, in *similar* contexts. Therefore, our distributional model indexes the semantic relationships between both words which co-occur in the same texts and paragraphs (first-order co-occurrence: for example, words linked to a common event or function, e.g. Nelson, 1983, Lin and Murphy, 2001; Crutch and Warrington, 2005, e.g. *peach* and *knife*, *peach* and *to peel*, *harp* and *to play*), and, most importantly, words which do not appear in the same text, but can co-occur in similar

contexts (second order co-occurrence). For instance, although *play* and *drum* may not appear in the same text, they may separately co-occur with words like *music hall*, *sound* or *hands*. Thus, the distributional model captures abstract, second-order semantic information about word meanings, reflecting the statistical knowledge about their actual usage in language. We applied Word2Vec to the British National Corpus (BNC), which includes 4096 texts with samples of written and spoken English from a wide range of sources and a variety of genres for a total of 100 million words. Semantic similarity between words for each condition was then measured as the cosine between two-word vectors: the smaller the cosine, the greater the similarity between word stimuli pairs. Figure 1 (top left panel) displays the semantic similarity values based on Word2Vec in the form of RDMs. The corresponding arrangement of the experimental words was also visualised using multidimensional scaling via MDS graphs (Fig. 1 left bottom panel).

*Action-relatedness Model.* For a more precise interpretation of the link between the action semantics inherent in the nature of our stimulus words and the word-elicited response patterns on the neurocognitive level, we included a model coding for rated action-relatedness. This model was based on the behavioural data from an independent rating study (see Table 1 and Fig. 3, left panel). We hypothesized that the “action-relatedness” model explained the brain activity patterns in category-preferential regions supporting action semantics, especially in pMTG and motor cortex (e.g. Hauk et al., 2008; Carota et al., 2017), which were also expected to index taxonomic similarities (see Introduction).

*Control Model.* An additional model accounted for the familiarity of the words, a variable which is highly correlated with word frequency (controlled in the present study), and can affect word-

elicited brain response patterns in frontal and temporal regions during word recognition (e.g. Hauk et al., 2008). The model was expected to correlate with similarity patterns in regions which were also expected to reflect co-occurrence frequency of the words, such as the LIFG. The model was constructed using behavioural ratings of familiarity (see Table 1). The corresponding RDM is shown in Fig 2 (right panel).



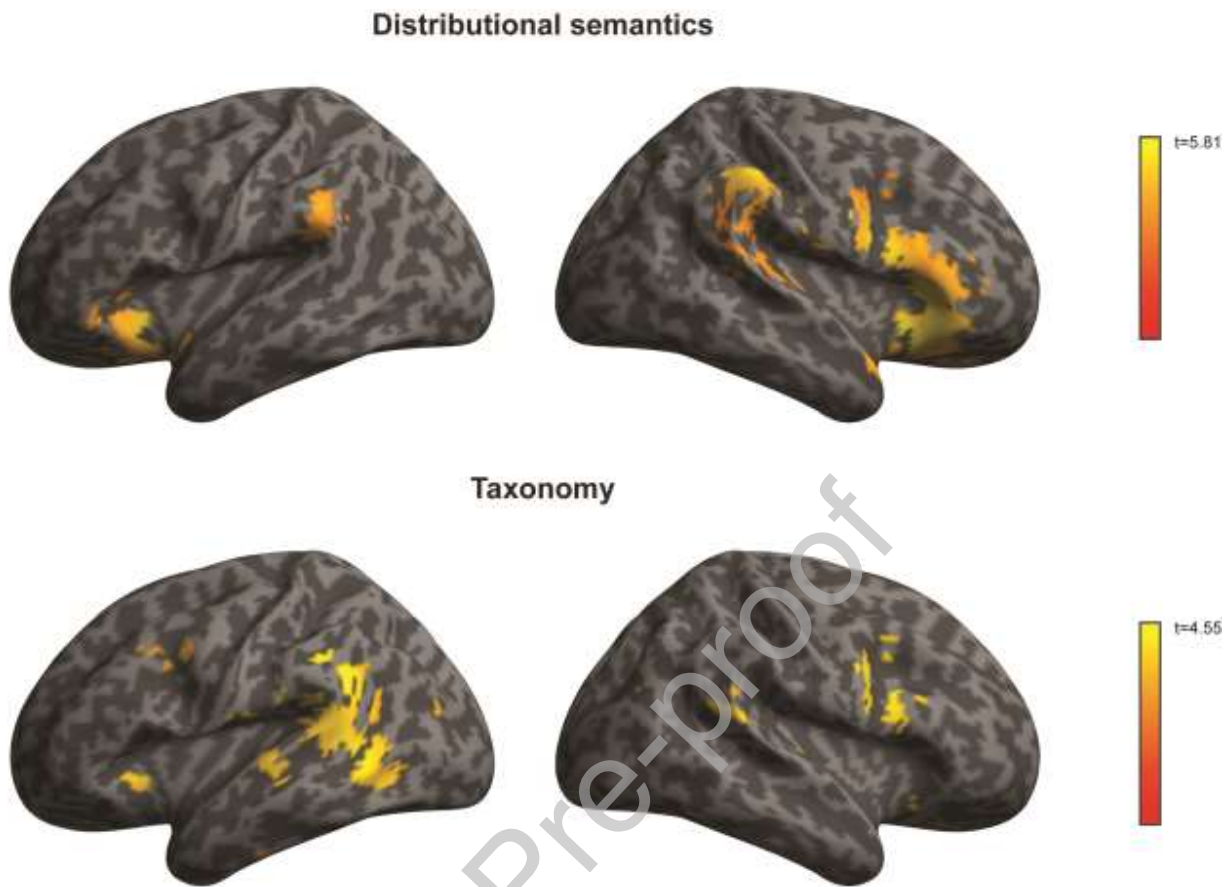
**Figure 2.** The representational dissimilarity matrices (RDMs: 1 = dissimilarity, in yellow -, 0 = similarity, in blue) for the models expressing Action-relatedness (left panel) and Familiarity (right panel). Note the difference in the representational geometries predicted by the two models. In the Action-relatedness model, words related to arm (grey), face (red), and leg (blue) were represented as being similar (left top blue square of the RDM). Object words were coded as being similar to each other (right bottom blue square of the RDM), but differed from action words, with the exception of the tools, which, in turn, were represented as being similar to actions. In contrast, the Familiarity model did not exhibit any neat categorical structure.

## Results

*Results from whole-brain searchlights: taxonomic and distributional models were reflected in distinct fronto-temporal regions.*

Results from whole-brain searchlights showed that the distributional model correlated with activity patterns in a network of fronto-temporal and inferior parietal regions, which have been found to be relevant to semantic comprehension across studies and tasks (e.g. Binder et al., 2009; see discussion). As depicted in Table 2 and Fig. 4 (top panel), the Word2Vec model matched the similarity of fMRI patterns in pars orbitalis of the IFG (BA 47) and SMG/AG bilaterally. All effects survived cross-validation.

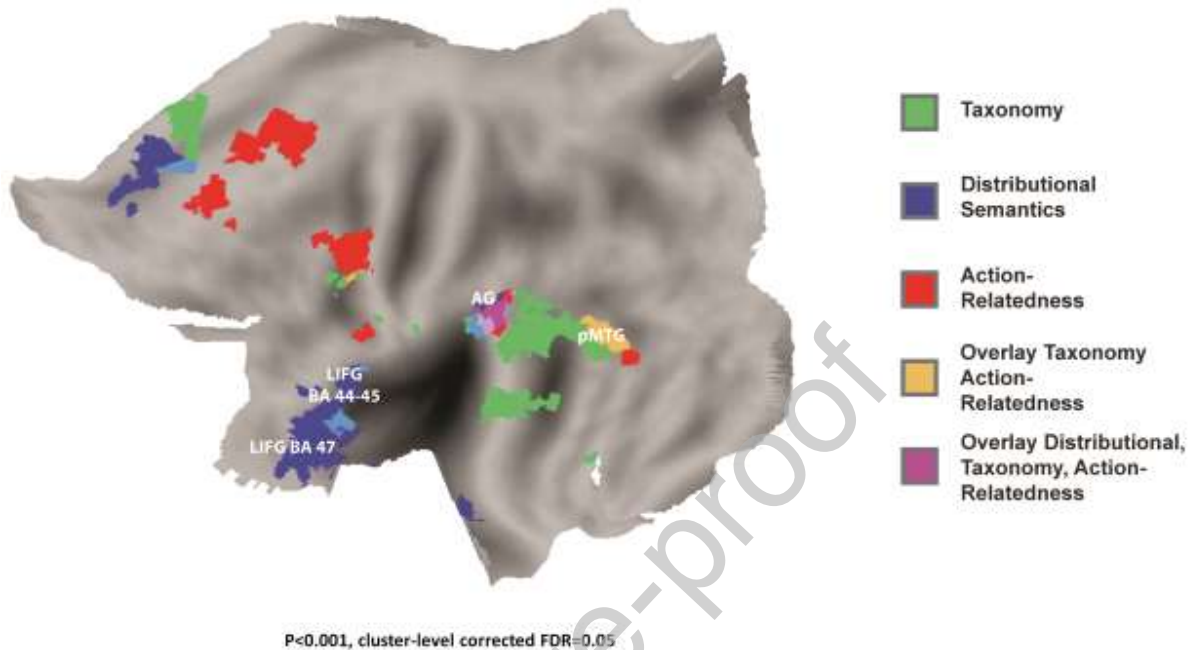
In contrast, the taxonomy-based model triggered a focused effect in left posterior temporal cortex, peaking in pMTG, a region important for the storage of lexico-semantic representations in long-term memory, as consistently shown across different MVPA studies (Fairhall and Caramazza 2013; Devereux et al. 2013; Clarke and Tyler, 2014; Carota et al., 2017) (see Table 3 and Fig. 4, bottom panel, and Table 2 and Fig. 1 in Inline Supplementary Materials). Additional model correlations were seen in the pars opercularis of the LIFG (BA 44) and precentral gyrus (henceforth PG) of the motor cortex.



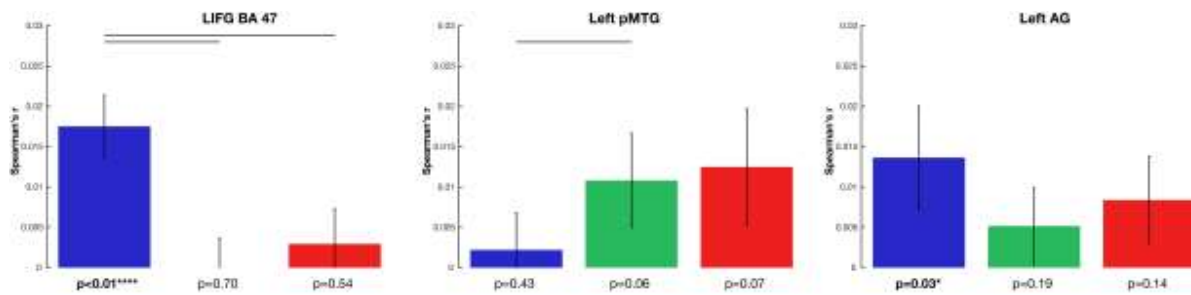
**Figure 3.** Significant effects in RSA searchlights ( $p_{\text{FWE}} < 0.05$ ), following cross-validation. Top panel: results from whole-brain searchlight RSA showing the significant correlations of the distributinal model with the similarity patterns in the pars orbitalis of the LIFG (BA 47) and in the AG. Bottom panel: results from whole-brain searchlight RSA showing the significant correlations ( $p_{\text{FWE}} < 0.05$ ) of the taxonomic model in left posterior middle temporal and inferior parietal regions, and motor cortex.

Turning to the effects of rated action-relatedness, we found that the action model correlated with the similarity structure of the response patterns in left pMTG, LIFG (BA 44) and adjacent PG, and AG (see Table 4). As displayed in Fig. 4, the taxonomic model triggered

correlations with the response patterns in these same regions; in left pMTG and motor regions, these patterns also overlapped with the one specific to the extended action category model.



**Figure 4.** Significant effects in RSA searchlights for action-relatedness (in red), distributional semantics (in blue), and taxonomy (in green). The neural overlap between the three models in the AG is shown in purple. The activation patterns for action-relatedness overlapped with the pattern specific to taxonomy only in the left pMTG (overlap is in dark yellow). Results are shown at a threshold of  $p < 0.001$ , with cluster-level correction for  $FDR = 0.05$ .



**Figure 5.** Results from model comparisons. The bar graph depicts the averaged model-fMRI pattern correlations for each of the key variables under examination. In each panel, the blue bar indicates the distributional model, the green bar the taxonomic model and the red bar the action-related model. Spearman's rank correlations were calculated to assess the relatedness between brain activity and model RDMs and statistical inference was applied on the single subject correlations using a one-sided signed-rank test across subjects, testing whether the resulting correlation coefficients were significantly greater than zero. Below each bar, the significance value for the test is reported, corrected for multiple testing across brain regions by applying the FDR procedure; the expected FDR was less than 5% (Benjamini and Hochberg 1995). The horizontal bars in black indicate significant differences from model comparisons after FDR correction across models (FDR=0.05). Left panel: Effects specific to the LIFG (BA 47), where the distributional model differed significantly with both taxonomy and action-relatedness. Middle panel: Effects specific to the left pMTG. There was a significant difference (FDR=0.05) between the effects of the taxonomic model and the ones triggered by the distributional model, whilst no difference with action-relatedness was seen. Right panel: effects of the distributional, taxonomic, and action-relatedness in the AG. Note that all models were equally represented in this region.

As for the ratings of familiarity, the corresponding model correlated significantly with the neural patterns in the LIFG (BA 44), insula, lingual gyrus, and right MTG. None of these effects overlapped with the ones that were specific to the distributional and taxonomic semantic models (see Table 5 and Inline Supplementary Fig 2).

***Results from model comparisons: taxonomic and distributional representations were predominant in distinct fronto-temporal regions.***

Direct statistical comparisons between the distributional and taxonomic model in pre-selected ROIs revealed a statistically robust differentiation between the underlying representations in two focused regions. The distributional model (Word2Vec) showed significantly stronger correlations (FDR=0.05) than the taxonomic model (WordNet) with the patterns in the pars orbitalis of the LIFG (BA 47). Reversely, we found a better performance of the taxonomic model relative to the distributional one (FDR=0.05) in the left pMTG. These results indicate significant model specificity for these two regions, further confirming the searchlight data (see Fig. 5). The AG showed fair fits with both semantic models.

***Results from data cross-validation***

Cross-validated results from LDC searchlights suggested that the dissimilarity patterns related to the semantic models were highly stable and reliable in two regions, namely the pars orbitalis of the LIFG (BA 47) and the left pMTG, which respectively survived data cross-validation specific to co-occurrence (Table 2) and taxonomy (Table 3).



Additional patterns surviving cross-validation were seen in the pMTG and motor cortex for action-relatedness (see Table 4). In turn, familiarity correlated with the neural patterns in left insula and LIFG (BA 44), and occipital cortex (Table 5).

## Discussion

In the present study, we adopted a word reading non-semantic task to explore how distributional and hierarchical/taxonomic links among concepts relate to the semantic content of the language system. We tested the hypothesis that the distributional models would correlate with the neural patterns in inferior frontal and inferior parietal regions supporting combinatorial semantic processes, whilst the hierarchical/taxonomic model would fit the neural patterns in posterior middle/inferior temporal regions relevant for the comprehension of action words and action-related object words, such as tools and foods. A related hypothesis was that the hierarchical/taxonomic relations would relate to the neural patterns in posterior temporal and motor regions that were also sensitive to action-relatedness. Therefore, we also expected that the encoding of hierarchical/taxonomic relations would be consistent with the action-specific motor detail of our set of action-related words, as predicted by a model of behaviourally rated action-relatedness of the words in these category-specific fronto-temporal and motor regions.

Converging cross-validated results from combined RSA searchlights and ROI-based statistical model comparisons suggested that models of the different semantic relationships, the taxonomic model employing conceptual hierarchies and the distributional model built from word co-occurrences in texts, explained the similarity structure of region-specific fMRI response

patterns elicited by word reading. We found that the taxonomic model correlated with activity patterns in posterior middle temporal regions (pMTG) important for semantic memory (e.g., Fuster, 1997; Hagoort et al., 2014). In contrast, the distributional semantic model correlated with activity in the pars orbitalis of the LIFG (BA 47), a region thought to be particularly relevant for semantic selection (Thompson-Schill et al., 1997), and combinatorial semantic processes (Pulvermüller 2013; 2018). Confirming our initial hypotheses, in the same left posterior temporal regions which reflected taxonomy - but not in the LIFG (BA 47) -, the similarity patterns also conformed to the similarity expressed by the behaviourally rated action-relatedness of the words. In a third semantic region, the AG, the fMRI patterns equally reflected the distributional, taxonomic, and action-related models. It is worth noticing that the neural patterns in the LIFG (BA 47), pMTG, and AG, which correlated with the distributional and taxonomic models, did not relate to familiarity (as a particular aspect of word frequency, which was controlled for in the present study). These results suggest that the brain responses to lexical meaning reflect qualitatively distinct types of semantic relations among concepts within a distributed network of brain regions with different cortical distributions. These findings suggest a joint role of these regions in orchestrating the representational content of lexical meanings and their relationships, in line with their specific contributions to efficient semantic comprehension, as discussed in the next paragraphs. Although the effects we observed were to some extent bilaterally distributed, the next section will discuss them in more detail, focusing on the components identified in the left language regions.

*Neural overlaps between taxonomic and action-related similarities in pMTG and motor regions*

A major finding of the present study highlights the importance of hierarchical/taxonomic knowledge for representing lexical semantics in posterior middle temporal regions. These regions are thought to be a critical site for representation and processing of lexical meaning (e.g. Lau et al., 2008; Hickock, 2014), and play a fundamental role in distributed memory circuits supporting both language comprehension and production (e.g. Hagoort, 2014, 2019). It is well established that the pMTG is responsible for long-term storage and activation of semantic knowledge in memory (e.g., Fuster, 1997; 2009; Hagoort, 2019), with a possible role in controlled semantic retrieval requiring the extraction of information appropriate for a given context. For example, recent results have shown that semantic judgements about global relations between concepts like “apple” and “worm”, also called weak semantic or thematic associations, activate the pMTG more strongly than judgments comparing the size of two objects (Davey et al., 2015 a). Earlier neuropsychological findings also reported that lesions to this region cause weak semantic associative errors in aphasia (Schwartz et al., 2011). Our present results corroborate a role of the pMTG in this type of relational semantics, as the taxonomic model captured cross-category referential links between action and action-related object words (e.g. tools), by further specifying the nature of the underlying information patterns and their representational geometries. It is noteworthy that, in the present data, the similarity structure of the hierarchical/taxonomic model in pMTG incorporated both within-category similarities between conceptually related pairs like “plum” and “peach” and weak semantic associations between thematically linkable (i.e. weak, or global semantically associated) word pairs like “peach” and “bite”. Given such cross-category relations in the taxonomic model, our data line up with previous evidence that the pMTG encodes weak semantic associations, also labelled as thematic relations by earlier work (Schwartz et al., 2011). As mentioned earlier (see

Introduction), the hierarchical/taxonomic relations expressed by our WordNet model capture referential “word-world” links among action and object concepts, reflecting extra-linguistic, world knowledge. Although such links may then coincide with weak semantic, or thematic associations between words that may occur together in the same context, they differ crucially from the second-order co-occurrence links expressed here by distributional semantics, which bases its computational power on the capacity to generalise semantic information to items that may co-occur in similar contexts, and primarily reflect context-sensitive word-word relations.

A related finding was indeed the congruency between the similarity in action-relatedness and the representational geometries of the response patterns in the pMTG. This effect was seen in the same posterior middle temporal regions, for which seminal univariate results have reported correlation of BOLD activation to action words with their rated action-relatedness (Hauk et al., 2008). Our present results then reinforce the well-established view that the pMTG supports category-preferential representations of action semantics (Chao and Martin 1999; Kiefer and Pulvermüller 2012; Beauchamp and Martin 2006; Noppeney et al. 2005; Saygin et al., 2009; Tranel et al. 2003; 2005; Hoenig et al. 2011; Kemmerer et al. 2012; Davey et al., 2015 b). Likewise, the similarity in action-relatedness correlated with the representational geometries in left PG, a seat for action semantic mechanisms (Tranel et al., 2008; Hauk et al., 2004; Carota et al., 2012, 2017; Pulvermüller, 2013; Vukovic and Shtyrov, 2019), as well as in adjacent dorsal aspect of the LIFG (BA 44) (Bak and Chandran 2012; Kemmerer et al., 2012; Kemmerer, 2015; Dreyer et al. 2015), which earlier MVPA results suggested to reflect the semantic structure of action word categories (Carota et al., 2017). This suggests that the encoding of taxonomic similarities conforms, not surprisingly, to the grounded action-related information supported by

action processing systems (e.g., Barsalou, 2008; Pulvermüller and Fadiga, 2010; Kemmerer, 2015; Pulvermüller, 2018).

The taxonomic format of semantic representations stored in the pMTG along with its grounded sensorimotor specifications may be retrieved and made available for semantic processes taking place in the other language regions, such as the LIFG and inferior parietal cortex, as discussed in the next section.

### ***Distributional-specific cortical mapping in LIFG and AG***

Consistent with our initial hypothesis and earlier findings (Carota et al., 2016; 2017), the pars orbitalis of the LIFG (BA 47) emerged as a privileged region for representing semantic similarities based on statistical co-occurrence knowledge. This was not unexpected, since this area is thought to support lexical semantics (e.g. Poldrack et al., 1999), particularly in relation to the selection of the relevant semantic properties from competing semantic alternatives (Thompson-Schill, 1997), and plays a cross-linguistically validated function in semantic comprehension (e.g. Bozic et al., 2010; Carota et al., 2016).

The LIFG has been proposed to support category-general links between word meanings from all semantic types and word forms (Pulvermüller, 2013; Pulvermüller, 2018). In line with such proposal, a remarkable property of the representational patterns of the LIFG (BA 47) in the present results was the lack of any reflection of category-specific action-related structure. Note that these patterns were dissociable from the category-specific effects of action-relatedness in the pars opercularis of the LIFG (BA 44). In the LIFG (BA 47), the distributional representations seemed to rely on genuinely semantic processes on a higher level of abstraction. In particular, the

representation of abstract semantic information about the likelihood of words to co-occur in language contexts may be a prerequisite for the semantic combinatoric mechanisms taking place in this region. Availability of the context-sensitive semantic representations that can fit similar semantic contexts may be particularly important for unification mechanisms which aim at assembling semantic information into larger structured units of coherent sentences and discourse and culminate in left inferior frontal cortex (e.g. Hagoort, 2005).

The present results thus confirm a joint contribution of the pMTG and more anterior aspect of the LIFG for activation, maintenance and control of retrieved lexical meanings in language comprehension (e.g., Lau et al., 2008; Snijders et al., 2009; Turken and Dronkers, 2011; Hickock, 2014). In the context of sentence comprehension, for instance, it has been shown that increased ambiguity of word-category sequences (relative to unambiguous sequences: Snijders et al., 2009, 2010) produces more sustained activation of long-term representations of lexical frame information in the pMTG, controlled top-down by inferior frontal regions (Fuster 2001; Miller and D'Esposito 2005). Furthermore, relevant neurophysiological results obtained with methods with a more precise temporal resolution than what fMRI provides have pointed out that the MTG may give input to the critical language regions, such as inferior frontal cortex (BA 45-47), and inferior parietal cortex, particularly the AG (e.g. Lau et al., 2008), thus enabling access to activated lexical-semantic representations. Our results further elucidate the nature and distribution of lexical-semantic representations in this cortical network.

We can now turn to a third major effect of semantic similarities, which arose indeed in the AG, an associative convergence region essential for the multimodal semantic integration of unimodal inputs received from unimodal sensorimotor systems (e.g. Binder et al., 2009; Binder

and Desai, 2011; Geschwind, 1967; Catani et al., 2005). Recent evidence has shown that the AG supports the comprehension of contextually appropriate complex concepts (e.g. plaid jacket) based on simple conceptual units (e.g. jacket and plaid) (Price et al., 2015). Lesions to this region impair association-based representations in aphasic patients, as indexed by associative naming errors (Schwartz et al., 2011). Also, earlier metabolic studies have demonstrated that, in the AG, conceptual similarities may arise from the multimodal integration and abstraction of sensorimotor attributes (Fernandino et al., 2015). Whilst our present results fit this previous literature well, confirming a role of the AG in relational semantics, they reveal a threefold format of the representational similarities in this region, reflecting hierarchical/taxonomic, context-sensitive distributional, and action-related representations. A plausible explanation is that the AG binds together the referential knowledge inherent to the taxonomic similarity and the statistical information about word co-occurrence in coherence with the appropriate sensorimotor detail. The reflection of rated action-relatedness in the AG, also spanning to the more posterior sections of the SMG, reflects, once more, the action-related nature of our stimulus words (action words and action-related manipulable foods and tools), consistent with earlier findings on the role of these regions in action semantics (e.g., Davey et al., 2015 a), and in the representation of manipulable objects (e.g. Cubelli et al. 2000; Buxbaum and Saffran 2002; Mahon et al. 2007; Martin et al. 2015).

A last observation concerns the absence of effects of categorical representations in the AG. In earlier work, we investigated the categorical semantic structure of the neural patterns elicited by specific semantic word categories (Carota et al., 2017), and found that pMTG, motor and inferior frontal regions were important to encode lexical-semantic similarities, but observed no effects in the AG. We therefore raised the possibility that this region may be more sensitive to

gradual representational differences among individual word meanings, rather than categorical representations, and therefore be specifically linked to category-general encoding of finer-grained semantic representations of individual word links. Indeed, the use of sensitive item-by-item RSA analyses allowed us to identify specific types of semantic similarities in the inferior parietal cortex.

### *Semantic specificity vs. task specificity of semantic similarity mapping*

To evaluate the cognitive implications of our present results for current theories of lexical-semantic processing, it is worth noticing that task-specific processing demands may have differently modulated the neural activity in the network of cortical regions we identified here, leading to increased interactions with additional semantic regions, such as the aTL, which was not detected in the present study. This may explain discrepancies between the present findings and the recent results. For example, valuable RSA evidence suggested that overt similarity judgments of taxonomic vs. thematic similarity relations triggered dissociable activation patterns in the left aTL and in the AG, respectively (Xu et al., 2016), leading to the conclusion that taxonomic knowledge may be a primary form of semantic representation in the aTL (Xu et al., 2016).

One obvious reason for the lack of effects in the aTL across neuroimaging studies is the widely acknowledged fMRI signal dropout and distortion issues due to varying magnetic susceptibility (e.g., Devlin et al., 2000, Visser et al., 2010). In the present data though, the univariate subtractive contrast comparing all visually presented words against the low-level visual baseline showed activation in a region located in the aTL, which may however index different processes, ranging from letter recognition to word form and phonological processing,



semantic comprehension and even processing of aspects of the syntactic information associated with the words. Therefore, task specificity seems to offer an appropriate methodological explanation here, as neuroimaging studies consistently show that BOLD activation in this region increases for semantic tasks relative to non-semantic tasks (e.g., Visser et al., 2010). This has also been confirmed across imaging modalities, as previous electro-/magneto-encephalographic data have pointed to increased activity in tasks involving semantic decision relative to silent reading (Chen et al., 2013). As mentioned above, we here employed a non-semantic task, because it may tap into automatic processes of symbol meaning recognition, which are not required for task-dependent performance, and, therefore, allows to draw conclusions on language comprehension mechanisms. In particular, models of word recognition posit that activating a lexical representation is a necessary and sufficient stepping stone for accessing its meaning (Balota and Yap, 2006), as, for instance, lexical identification is necessary for the execution of behavioural responses such as lexical decision. Therefore, activating lexical-semantic representations is a flexible process, which adapts to task requirements, and is modulated by attentional control and task-dependent demands (Balota and Yap, 2006). The present experiment tapped into such automatic and flexible activation of lexical semantic representations taking place by viewing words. That the attentive reading of meaningful words yielded representational similarities reflecting both complexity and specificity of semantic components thus confirms semantic understanding of symbols despite the absence of a semantic task, a finding consistent with earlier research (see, for example, Hauk et al., 2004; Carota et al., 2012 and, for reviews, Binder and Desai, 2011). Furthermore, the RSA manifestation of semantic similarity in a network of regions that are critical for language comprehension is evidence that word meaning was processed (Carota et al., 2017).

If the absence of a semantic task explains the general lack of specific multivariate effects of semantic similarities in anterior temporal regions in the present study, it is plausible that, unlike automatic meaning processing, actively and overtly judging the similarity of words based on taxonomic relations may increase the load in processing demands in favour of the corresponding information type in the aTL. Relatedly, a possibility is that the aTL is not engaged by semantic tasks which involve semantic coding per se, but is required for connecting brain regions that primarily store lexical semantics (e.g.; Damasio et al., 2004).

In conclusion, the present results are not incompatible with the suggestion that the aTL may become representationally relevant for encoding taxonomic and associative relations, depending on task requirements that may put into play different processing demands in the language networks, and modulate the sensitivity of this region to a certain type of semantic links. However, our findings did not bring evidence for a single representational hub located in the aTL and, rather, point to distributed representational systems across multiple fronto-temporal and parietal regions (e.g.: Martin and Chao, 2001; Binder et al., 2009; 2016; Martin, 2016; Pulvermüller, 2018; Pereira et al., 2018).

### *Neurocognitive relevance of the present findings for current models of semantic representations*

We outlined a neurobiological model of language comprehension which, captures two substantial components of the representation of lexical meaning: 1) its relational nature and 2) the distribution of its representations between key regions of the language systems. These included temporo-parietal regions (pMTG and AG) encoding taxonomic similarity consistent

with the sensorimotor properties of the word semantic types, and the pars orbitalis of the LIFG (BA 47) encoding distributional statistics relevant for semantic combinatorics and unification of lexical-semantic units into larger structures as contextually appropriate. This network also exhibited a gradient, from category-specific encoding of action-related semantic information in pMTG to category-general effects in the AG and LIFG (BA 47), where the representational content of words became, respectively, more sensitive to individual concepts along with their action-related details and to general statistical word-word relations than to category-specific representations of semantic word structure. The present results match well integrated theories that acknowledge a task-sharing between multimodal modality-specific areas, whereby multiple multimodal regions play a main role in binding referential meanings (i.e., links between word forms and their referent concepts, e.g., actions and objects, de Saussure, 1860; see above and introduction), grounded by neuronal assemblies in modality-specific cortex, with distributional semantic knowledge (Barsalou, 2008; Harnad, 2012; Kiefer and Pulvermüller, 2012; Pulvermüller et al., 2018; Carota et al., 2017).

The distributed representations are well explained by widespread networks of interconnected cell assemblies (e.g., Hebb 1949), functional units of memory defined by neural relationships (Fuster, 1997). These circuits carry semantic information with specific distributions and distinct topographies depending on word types, and are scattered over both category-general regions in supramodal associative cortex and category-preferential areas in modality-specific cortex (e.g. Pulvermüller, 2013; 2018).

While confirming the involvement of a left inferior frontal, and temporo-parietal regions in word meaning comprehension, the current results highlight both taxonomic and word distributional knowledge as relevant representational coordinates for characterizing the

underlying semantic space. Taxonomy may provide a format for semantic representations stored in left pMTG, and might well be a primary representational frame structuring conceptual knowledge in memory. In contrast, distributional knowledge could be stored in the LIFG (BA 47) as cached information immediately available to increase efficiency of operations like access and selection of "what comes next" during online comprehension and usage of the human vocabulary.

### **Conclusions**

In conclusion, we have shown that both hierarchical/taxonomic and distributional models of semantic similarities describe the representational space in a high-level network of inferior frontal and temporo-parietal regions known to play a critical role in semantic memory. These results support integrative semantic brain models postulating not a single, but cooperating representational systems, in which distinct brain regions may encode specific semantic properties of words to a different degree. The present study also demonstrates that RSA of BOLD signal offers an accurate tool for identifying the fine-grained representations of concepts and words. This and related methodological approaches (Haxby et al., 2014) allow investigation of the types of information that is coded in specific brain regions. Adjudicating between alternative representational models of word relationships combining standard RSA, cross-validation methods and model comparisons, as performed here, is a promising step to elucidate the nature of the semantic representations contributing to human language.

## References

Bak TH, Yancopoulou D, Nestor PJ, Xuereb JH, Spillantini MG, Pulvermüller F, Hodges JR. 2006. Clinical, imaging and pathological correlates of a hereditary deficit in verb and action processing. *Brain* 129:321-332.

Bak TH, Chandran S. 2012. What wires together dies together: verbs, actions and neurodegeneration in motor neuron disease. *Cortex* 48(7): 936-944.

Balota DA, Yap M. 2006. "Attentional control and flexible lexical processing: explorations of the magic moment of word recognition," in *From Inkmarks to Ideas: Current Issues in Lexical Processing*, ed S. Andrews (New York, NY: Psychology Press), 229–258.

Barsalou LW. Grounded cognition. 2008. *Annu Rev Psychol* 59: 617-645.

Barsalou LW. 2017. What does semantic tiling of the cortex tell us about semantics? *Neuropsychologia* 105:18-38. (doi:10.1016/j.neuropsychologia.2017.04.011)

Beauchamp MS, Martin A. 2006. Grounding object concepts in perception and action: evidence from fMRI studies of tools. *Cortex* 43(3):461-468.

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao, SM, Prieto T. 1997. Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci* 17: 353-362.

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

Binder JR, Desai RH. 2011. The neurobiology of semantic memory. *Trends Cogn Sci* 15(11):527-36.

- Binder JR, Conant LL, Humphries CL, Fernandino L, Simons SB, Aguilar M, Desai RH. 2016. Toward a brain-based componential semantic representation. *Cognitive Neuropsychology* 33(3-4): 130-174.
- Boynton GM, Engel SA, Glover GH, Heeger DJ. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16(13): 4207-4221.
- Bookheimer SY. 2002. Functional MRI of Language: new approaches to understanding the cortical organization of semantic processing. *Annu Rev Neurosci* 25: 151-88.
- Bozic M, Tyler LK, Ives DT, Randall B, Marslen-Wilson W. 2010. Bihemispheric foundations for human speech comprehension. *107(40):17439-17444*.
- Buxbaum LJ, Saffran EM. 2002. Knowledge of object manipulation and object function: dissociations in apraxic and nonapraxic subjects. *Brain Lang* 82(2):179-199.
- Carlson TA, Simmons RA, Kriegeskorte N, Slevc LR. 2014. The emergence of semantic meaning in the ventral temporal pathway. *J Cogn Neurosci* 26:120-131.
- Carota F, Moseley R, Pulvermüller F. 2012. Body-part-specific Representations of Semantic Noun Categories. *J Cogn Neurosci* 24(6):1492-509.
- Carota F, Kriegeskorte N, Nili H, Pulvermüller F. 2017. Representational similarity mapping of distributional semantics in left inferior frontal, middle temporal and motor cortex. *Cereb Cortex* 27(1):294-309.
- Carota F, Bozic M, Marslen-Wilson W. 2016. Decompositional representation of morphological complexity: Multivariate fMRI evidence from Italian. *J Cogn Neurosci* 28(12): 1878-1896.
- Catani M, Jones DK, ffytche DH. 2005. Perisylvian language networks of the human brain. *Ann Neurol* 57(1):8-16.

- Chen Y, Davis MH, Pulvermüller F, Hauk O. 2013. Task modulation of brain responses in visual word recognition as studied using EEG/MEG and fMRI. *Front Hum Neurosci* 7:376. doi: 10.3389/fnhum.2013.00376.
- Clarke A, Tyler L. 2014. Object-Specific Semantic Coding in Human Perirhinal Cortex. *J Neurosci* 34(14):4766-4775.
- Cramer P. 1968. *Word Association*. New York: Academic Press.
- Crutch SJ, Warrington EK. 2005. Gradients of semantic relatedness and their contrasting explanations in refractory access and storage semantic impairments. *Cogn Neuropsychol* 22(7):851-76.
- Dapretto M, Bookheimer S. 1999. 'Form and content: Dissociating syntax and semantics in sentence comprehension'. *Neuron* 24: 427-432.
- Cubelli R, Marchetti C, Boscolo G, Della Sala S. 2000. Cognition in action: testing a model of limb apraxia. *Brain Cogn.* 44(2): 144-165.
- Damasio AR. 1989a. The Brain Binds Entities and Events by Multiregional Activation from Convergence Zones. *Neural Computation* 1(1): 123-132.
- Damasio AR. 1989b. Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33(1-2): 25-62.
- Damasio H, Grabowski TJ, Tranel D, Hichwa RD, Damasio AR. 1996. A neural basis for lexical retrieval. *Nature* 380(6574): 499-505.
- Damasio H, Tranel D, Grabowski T, Adolphs R, Damasio A. 2004. Neural systems behind word and concept retrieval. *Cognition* 92(1-2): 179-229.
- Davey J, Thompson HE, Hallam G, Karapanagiotidis T, Murphy C, De Caso I, Krieger-Redwood K, Bernhardt BC, Smallwood J, Jefferies E. 2015a. Exploring the role of the posterior

middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. *Neuroimage*: 137:165-177. doi: 10.1016/j.neuroimage.2016.05.051. Epub 2016 May 25.

Davey J, Rueschemeyer S-A, Costigan A, Murphy N, Krieger-Redwood K, Hallam G, Elizabeth Jefferies E. 2015b. Shared neural processes support semantic control and action understanding. *Brain and Language* 142: 24-35.

Demb JB, Desmond JE, Wagner AD, Vaidya CJ, Glover GH, Gabrieli JDE. 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. *J Neurosci* 33(48):18906-16.

Devlin JT, Russell RP, Davis MH, Price CJ, Wilson J, Moss HE, et al. 2000. Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. *Neuroimage* 11: 589-600.

Devlin JT, Matthews PM, Rushworth FS. 2003. Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *J Cogn Neurosci* 15: 71-84.

Dreyer FR, Frey D, Arana S, von Saldern S, Picht T, Vajkoczy P, Pulvermüller F. 2015. Is the motor system necessary for processing action and abstract emotion words? Evidence from focal brain lesions. *Front Psychol.* 6:1661.

Dronkers NF, Wilkins DP, Van Jr VR, Redfern BB, Jaeger JJ. 2004. Lesion analysis of the brain areas involved in language comprehension. *Cognition* 92: 145-177.



- Etzel JA, Zacks J, Braver TS. 2013. Searchlight analysis: Promise, pitfalls, and potential. *Neuroimage* (78): 261-269. DOI: [10.1016/j.neuroimage.2013.03.041](https://doi.org/10.1016/j.neuroimage.2013.03.041)
- Fairhall S, Caramazza A. 2013. Brain Regions That Represent Amodal Conceptual Knowledge. *J Neurosci* 33(25): 10552-10558.
- Fellbaum C. 1998. *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.
- 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. *Cereb Cortex* pii: bhv020.
- Firth JR. 1957. A Synopsis of Linguistic Theory, 1930-1955. In J.R. Firth et al., editors. *Studies in Linguistic Analysis*. Special volume of the Philological Society. Oxford: Blackwell.
- Fuster JM. 1997. Network memory. *Trends Neurosci* 20: 451-459.
- Fuster, JM. 2001. The prefrontal cortex – an update: time is of the essence. *Neuron* (30): 319-333.
- Fuster, JM. 2009. Cortex and memory: emergence of a new paradigm. *J Cogn Neurosci* 21: 2047-2072.
- Gabrieli JD, Poldrack RA, Desmond JE. 1998. The role of left prefrontal cortex in language and memory. *Proc Natl Acad Sci USA* 95: 906-913.
- Geschwind N. 1965. Disconnexion syndromes in animals and man. *Brain*. 88:237–294; 585-644.
- Glenberg AM, Mehta S. 2008. Constraint on covariation: It's not meaning. *Italian Journal of Linguistics*. 20(1): 241-264 24.
- Hagoort, P. 2005. On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9: 416-423.
- Hagoort, P. 2013. MUC (Memory, Unification, Control) and beyond. *Front Psychology* 4: 416.

- Hagoort, P. 2019. The meaning making mechanism(s) behind the eyes and between the ears. *Phil Trans R Soc B*. 375(1791): 20190301.
- Hanson SJ, Toshihiko M, Haxby JV. 2004. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: Is there a “face” area. *Neuroimage* 23:156-167.
- Harnad S. 2012. From Sensorimotor Categories and Pantomime to Grounded Symbols and Propositions. In M. Tallerman & K. R. Gibson (Eds.), *The Oxford handbook of language evolution* (pp. 387-392). Oxford: Oxford University Press.
- Harris Z. 1954. Distributional structure. *Word* 10(23): 146-162.
- Hauk O, Johnsrude I, Pulvermüller F. 2004. Somatotopic representation of action words in the motor and premotor cortex. *Neuron* 41: 301-307.
- Hauk O, Davis M, Pulvermüller F. 2008. Modulation of brain activity by multiple lexical and word form variables in visual word recognition: A parametric fMRI study. *Neuroimage* 2008 42(3):1185-95.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293: 2425-30.
- Haxby JV, Connolly AC, Guntupalli JS. 2014. Decoding neural representational spaces using multivariate pattern analysis. *Ann Rev Neurosci*: 37:435-456.
- Hebb D. O. *The Organization of Behaviour*. New York: Wiley. 1949.
- Hickok G. 2014. *The myth of mirror neurons: the real neuroscience of communication and cognition*. New York: WW Norton and Company.

- Hillis AE, Wityk RJ, Tuffiash E, Beauchamp NJ, Jacobs MA, Barker PB, Selnes OA. 2001. Hypoperfusion of Wernicke's area predicts severity of semantic deficit in acute stroke. *Ann Neurol.* 50(5):561-6.
- Hoenig K., Müller C., Herrnberger B., Spitzer M., Ehret G. and Kiefer M. 2011. Neuroplasticity of semantic maps for musical instruments in professional musicians. *NeuroImage* 56: 1714-1725.
- Jackson RL, Hoffman P, Pobric G, Lambon Ralph MA. 2015. The Nature and Neural Correlates of Semantic Association versus Conceptual Similarity. *Cereb Cortex* 25(11): 4319-4333.
- Kemmerer D, Rudrauf D, Manzel K, Tranel D. 2012. Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex* 48: 826-848.
- Kemmerer D. 2015. Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychol Bull Rev.* 22(4): 1068-1075.
- Kiefer M, Pulvermüller F. 2012. Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex* 48: 805-825.
- Kiani R, Esteky H, Mirpour K, Tanaka K. 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol.* 97(6): 4296-309.
- Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proc Natl Acad Sci U S A.* 103(10): 3863-8.
- Kriegeskorte N, Mur M, Bandettini PA. 2008. Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.
- Lambon Ralph, MA, Jefferies E, Patterson K, Rogers T. 2017. The neural and computational bases of semantic cognition. *Nat Rev Neurosci* 18: 42–55.

- Landauer TK, Dumais ST. 1997. A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychol Rev* 104: 211-240.
- Lau EF, Phillips C, Poeppel D. 2008. A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, 9, 920933.
- Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. 1997. Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging* 16(2), 187-198.
- Mahon BZ, Milleville SC, Negri GAL, Rumiati RI, Caramazza A, Martin A. 2007. Action-related properties shape object representations in the ventral stream. *Neuron*. 55(3):507–520.
- Mahon BZ, Caramazza A. 2010. Judging semantic similarity: An event-related fMRI study with auditory word stimuli. *Neurosci* 169: 279-286.
- Maki WS, Buchanan E. 2008. Latent structure in measures of associative, semantic, and thematic knowledge. *Psychonomic Bulletin & Review* 15 (3): 598-603.
- Martin A. 2007. The representation of object concepts in the brain. *Annu Rev Psychol* 58, 25-45.
- Martin M, Beume L, Kümmerer D, Schmidt CS, Bormann T, Dressing A, Ludwig VM, Umarova RM, Mader I, Rijntjes M, et al. 2015. Differential roles of ventral and dorsal streams for conceptual and production-related components of tool use in acute stroke patients. *Cereb Cortex*. 26(9):3754–3771.
- Martin A 2016. GRAPES-Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychon Bull Rev* 23: 979-90.
- 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.

- Meyer K, Damasio A. 2009. Convergence and divergence in a neural architecture for recognition and memory. *Trends Neurosci.* 32(7): 376-82.
- Mikolov T, Sutskever I, Chen K, Corrado GS, Den J. 2013. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems 26 (NIPS 2013)*.
- Miller GA. 1970. Empirical methods in the study of semantics. In Steinberg D., Jakobovits, editors. *Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology*. New York: Cambridge University Press; pp 569-585.
- Miller BT, D'Esposito M. 2005. Searching for "the top" in top-down control. *Neuron* 48, 535–538.
- Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X. 2004. Learning to decode cognitive states from brain images. *Mach. Learn.* 57: 145-17510.
- Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA. 2008. Predicting human brain activity associated with the meanings of nouns. *Science* 320: 1191-1195.
- Mitchell DJ, Cusack R. 2016. Semantic and Emotional Content of Imagined Representations in Human Occipitotemporal Cortex. *Scientific Reports* 6:20232. DOI: 10.1038/srep20232.
- Molinaro N, Paz-Alonso PM, Duñabeitia JA, Carreiras M. 2015. Combinatorial semantics strengthens angular-anterior temporal coupling. *Cortex* 65:113-27.
- Moseley R, Carota F, Hauk O, Pulvermüller F. 2015. A Role for the Motor System in Binding Abstract Emotional Meaning. *Cereb Cortex* 22(7): 1634-1647.
- Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N. 2014. A toolbox for representational similarity analysis. *PLoS Comp Biol* 10(4): e1003553.
- Noppeney U, Josephs O, Kiebel S, Friston KJ, Price CJ. 2005. Action selectivity in parietal and temporal cortex. *Cogn. Brain Res.* 25:641-49.

- Patterson K, Nestor PJ, Rogers TT. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci* 8: 976-987.
- Pereira F, Gershman S, Ritter S, Botvinick M. 2016. A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognitive Neuropsychology* 33(3-4): 175-190.
- Pereira F, Lou B, Pritchett B, Ritter S, Gershman SJ, Kanwisher N, Botvinick M, Fedorenko E. 2018. Toward a universal decoder of linguistic meaning from brain activation. *Nature Neurosci.* 9:963. DOI: 10.1038/s41467-018-03068-4
- Petersen SE, Fox PT, Snyder AZ, Raichle ME. 1990. Activation of extrastriate and frontal cortical areas by visual words and word- like stimuli. *Science.* 249: 1041-4.
- Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JDE. 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10: 15-35.
- Polyn SM, Natu VS, Cohen JD, Norman KA. 2005. Category-specific cortical activity precedes retrieval during memory search. *Science* 310(5756): 1963-6.
- Price AR, Bonner MF, Peelle JE, Grossman M. 2015. Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *J Neurosci* 35(7):3276-84.
- Pulvermüller F. 1999. Words in the brain's language. *Behav Brain Sci* 22: 253-336.
- Pulvermüller F, Kherif F, Hauk O, Mohr B, Nimmo-Smith I. 2009. Cortical cell assemblies for general lexical and category specific semantic processing as revealed by fMRI cluster analysis. *Hum Brain Mapp* 30: 3837-3850.
- Pulvermüller F, Fadiga L. 2010. Active perception: sensorimotor circuits as a cortical basis for language. *Nat Rev Neurosci.* 11:351-360.

- Pulvermüller F. 2013. How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn Sci* 17: 458-70.
- Pulvermüller F. 2018. Neural reuse of action perception circuits for language, concepts and communication. *Progress in Neurobiology* 60: 1-44.
- Quillian, M. R. 1968. *Semantic Memory, Semantic Information Processing*. In Minsky M. (Ed.). Cambridge, MS: The MIT Press, pp. 227-270.
- Sachs O, Weis S, Zellagui N, Sass K, Huber W, Zvyagintsev M, Mathiak K, Kircher T. 2011. How different types of conceptual relations modulate brain activation during semantic priming. *J Cogn Neurosci* 23(5):1263-73.
- Sass K, Sachs O, Krach S, Kircher T. 2009. Taxonomic and thematic categories: Neural correlates of categorization in an auditory-to-visual priming task using fMRI. *Brain Res* 1270:78-87.
- Saygin AP, McCullough S, Alac M, Emmorey K. 2009. Modulation of BOLD Response in Motion-sensitive Lateral Temporal Cortex by Real and Fictive Motion Sentences. *J Cogn Neurosci*. 22(11): 2480-2490.
- Schwartz MF, Kimberg DY, Walker GM, Brecher A, Faseyitan OK, Dell GS. et al. 2011. Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proc Natl Acad Sci*. 108: 8520-8524.
- Seghier ML, Fagan E, Price CJ. 2010. Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. *J Neurosci* 30(50): 16809-16817.
- Shiffrin RM. Memory search. In D. A. Norman (Ed.), *Models of memory*. New York: Academic Press. 1970: pp. 375-447.

- Snijders TM, Vosse T, Kempen G, Van Berkum JJA, Petersson KM, Hagoort P. 2009. Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cereb Cortex* 19: 1493-1503.
- Snijders TM, Petersson KM, Hagoort P. 2010. Effective connectivity of cortical and subcortical regions during unification of sentence structure. *NeuroImage* 52: 1633-1644.
- Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. 1997. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc Natl Acad Sci USA* 94: 14792-14797.
- Tranel D, Manzel K, Asp E, Kemmerer D. 2008. Naming dynamic and static actions: Neuropsychological evidence. *J Physiology* 102: 80-94.
- Trumpp NM, Kliese D, Hoenig K, Haarmeier T, Kiefer M. 2013. Losing the sound of concepts: damage to auditory association cortex impairs the processing of sound-related concepts. *Cortex* 49: 474-486.
- Turken AU, Dronkers NF. 2011. The neural architecture of the language comprehension network: Converging evidence from lesion and connectivity analyses. *Front Syst Neurosci* 201 (5):1.
- Visser M, Jefferies E, Lambon-Ralph MA. 2009. Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *J Cogn Neurosci* 22:1083-1094.
- Vukovic, N, Shtyrov, Y. 2019. Learning with the wave of the hand: Kinematic and TMS evidence of primary motor cortex role in category-specific encoding of word meaning. *Neuroimage* (202): 116179.
- Walther A, Nili H, Ejaz N, Alink A, Kriegeskorte N, Diedrichsen J. 2015. Reliability of dissimilarity measures for multi-voxel pattern analyses. *NeuroImage* 137: 188-200.



Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. 2001. Separate neural subsystems within 'Wernicke's area'. *Brain* 124: 83-95.

Xu Y, Wang X, Wang X, Men W, Gao J-H, Bi Y. 2018. Doctor, Teacher, and Stethoscope: Neural Representation of Different Types of Semantic Relations. *Journal Neurosci* 38(13): 3303-331.

Journal Pre-proof

## Acknowledgments

Fund: Medical Research Council (MC\_US\_A060\_0034, U10055.04.003.00001.01 to FP and MC-A060-5PR20 to NK) and European Research Council Starting Grant (ERC-2010-StG 261352 to NK). *Conflict of Interest*: None declared.

## Tables

**Table 1. Psycholinguistic properties and semantic ratings are shown for each word category.**

	<i>Arm verbs</i>	<i>Leg verbs</i>	<i>Face verbs</i>	<i>Animal nouns</i>	<i>Food nouns</i>	<i>Tool nouns</i>	Main effect of word-type (F)
<b>Length</b>	4.73 (.16)	4.6 (.15)	4.53 (.12)	4.73 (.15)	4.60 (.12)	4.53 (.15)	.533 (p=.75)
<b>Bigram freq.</b>	31248.6 (3138.6)	32472.14 (4035.5)	29029.4 (11551.4)	31538.82 (2985.33)	32699.44 (3939.16)	30550.2 (3422.14)	.155 (p=.98) .401 (p=.85)
<b>Trigram freq.</b>	2475.5 (283.6)	2673.85 (367.8)	2535 (301.52)	2601.56 (1598.1)	2825.27 (1199.01)	2209.2 (1200.5)	
<b>No. of neighbours</b>	5.2 (.9)	5.06 (.6)	5.437 (0.84)	5.2 (.76)	4.8 (.78)	5.25(.84)	.071 (p=.99)
<b>No. of meanings</b>	1.06 (.06)	1.125 (.34)	1.125 (.08)	1.125 (.08)	1.11 (.06)	1.187 (.10)	.333 (p=.89)
<b>Log. word freq.</b>	.66 (.11)	.58 (.08)	.57 (.12)	.61 (.15)	.60 (.12)	.71 (.15)	.289 (p=.91)
<b>Imageability</b>	4.47 (0.17)	4.43 (0.29)	3.97 (0.26)	6.32 (0.09)	5.48 (0.27)	5.35 (0.35)	11.85 (p<.0001)
<b>Concreteness</b>	4.14 (0.16)	3.59 (0.19)	3.62 (0.19)	6.60 (0.08)	6.21 (0.19)	5.73 (0.21)	60.67 (p<.0001)
<b>Action-relatedness</b>	4.83 (0.23)	4.88 (0.26)	5.31 (0.25)	1.60 (0.10)	2.02 (0.29)	3.22 (0.41)	34.18 (p<.0001)
<b>Face-relatedness</b>	1.56 (0.01)	1.39 (0.09)	5.75 (0.23)	1.20 (0.07)	2.06 (0.26)	1.27 (0.10)	122.02 (p<.0001)
<b>Arm-relatedness</b>	5.68 (0.13)	1.81 (0.14)	1.33 (0.09)	1.11 (0.05)	1.37 (0.12)	2.85 (0.36)	94.88 (p<.0001)
<b>Body Sensation</b>	3.74 (0.29)	3.49 (0.23)	3.92 (0.31)	1.16 (0.07)	1.35 (0.11)	1.40 (0.16)	39.65 (p<.0001)
<b>Valence</b>	3.45 (0.27)	4.05 (0.21)	3.66 (0.31)	3.52 (0.09)	4.08 (0.13)	3.85 (0.14)	1.67 (p=0.14)
<b>Arousal</b>	3.25 (0.27)	3.01 (0.21)	2.60 (0.28)	1.30 (0.21)	1.44 (0.16)	1.65 (0.19)	14.45 (p<.0001)
<b>Familiarity</b>	4.52 (0.17)	4.27 (0.14)	4.59 (0.25)	4.83 (0.22)	5.28 (0.33)	5.02 (0.26)	2.42 (p=.04)

**Table 2. Results from RSA searchlights for the distributional model. Table of coordinates and significance voxel-level peak values (p) in each activation cluster that was correlated with the Word2Vec model. Regions for which the activation patterns survived the LDC cross-validation are marked in bold.**

Regions	Cluster Extent	Voxel-level P	Pseudo T	Coordinates		
				x	y	z
<b>Word2Vec-model</b>						
<b>Left Inferior Frontal Gyrus Orbitalis (BA 47)</b>	350	0.0044	5.21	-33	23	-9
Left Superior Frontal Gyrus	122	0.0055	5.05	-24	17	55
<b>Left Angular Gyrus (BA 39)</b>	120	0.0090	3.90	-39	-66	36
Left Supramarginal Gyrus (BA 40)	54	0.0073	4.32	-57	-40	29
<b>Right Insula</b>	970	0.0044	5.84	29	20	-9
<b>Right Inferior Frontal Gyrus Orbitalis (BA 47)</b>		0.0044	5.31	34	23	-12
Right Superior Temporal Pole		0.0044	5.31	51	17	-20
<b>Right Supramarginal Gyrus</b>	660	0.0044	5.11	51	-28	44
<b>Right Superior Parietal Cortex</b>		0.0044	4.89	21	-58	55
Right Superior Frontal Gyrus	137	0.0055	3.93	21	14	55
Right Supplementary Motor Area		0.0073	3.87	6	29	-20
Right Middle Frontal Gyrus		0.0073	3.69	27	20	48
Right Superior Frontal Gyrus	42	0.0055	3.83	30	-7	59
Right Middle Frontal Gyrus		0.0080	3.23	42	-7	55

**Table 3. Results from RSA searchlights for the taxonomic model.** Table of coordinates and significance voxel-level peak values ( $p$ ) in each activation cluster that was correlated with similarity in rated WordNet synsets. Regions for which the activation patterns survived the LDC cross-validation are marked in bold.

Regions	Cluster Extent	Voxel-level P	Pseudo T	Coordinates		
				x	y	z
<b>Behaviourally validated WordNet</b>						
Left Inferior Frontal Gyrus Opercularis (BA 44)	65	0.0091	4.29	-51	8	14
Left Insula		0.0147	3.73	-33	17	10
Left Precentral Gyrus		0.0109	3.16	-36	5	36
Left Supramarginal Gyrus (BA 40)	380	0.0098	4.28	-48	-46	25
Left Angular Gyrus (BA 39)		0.0098	3.91	-37	-60	36
<b>Left Middle Temporal Gyrus</b>	490	0.0031	4.44	-54	-61	-1
Left Supplementary Motor Area	190	0.0073	4.81	-3	8	66
Left Supplementary Motor Area		0.0073	4.32	-6	17	62
Right Superior Medial Frontal Gyrus	185	0.0079	4.57	3	50	32
Left Anterior Cingulate Cortex		0.0079	3.96	-6	50	14
Right Precentral Gyrus	174	0.0080	4.65	51	-10	51
Right Postcentral Gyrus		0.0080	4.44	54	-25	48
Right Superior Frontal Gyrus	133	0.0080	4.97	24	17	55
Right Middle Occipital Gyrus	27	0.0080	4.33	42	-76	21

**Table 4. Results from RSA searchlights for the Action-relatedness model.** Table of coordinates and significance voxel-level peak values ( $p$ ) in each activation cluster that was correlated with rated action-related similarity among the test words. Regions for which the activation patterns survived the LDC cross-validation are marked in bold.

Regions	Cluster Extent	Voxel-level $P$	Pseudo $T$	Coordinates		
				$x$	$y$	$z$
<b>Action-relatedness</b>						
<b>Left Supplementary Motor Area</b>	500	0.0036	2.95	-3	20	48
<b>Left Precentral Gyrus</b>	180	0.0081	2.75	-42	2	51
Left Supramarginal Gyrus	140	0.0064	2.57	-57	-43	29
<b>Left Middle Temporal Gyrus</b>	90	0.0087	2.44	-48	-61	-1
Right Inferior Frontal Gyrus (BA 45)	150	0.0055	3.05	51	29	14
Right Superior Frontal Gyrus	340	0.0025	3.51	51	8	29
<b>Right Precentral Gyrus</b>	600	0.0010	3.50	24	17	55
		0.0088	2.75	51	-10	48
Right Supplementary Motor Area	400	0.0010	3.31	3	47	40

**Table 5. Results from RSA searchlights for familiarity.** Table of coordinates and significance voxel-level peak values ( $p$ ) in each activation cluster that was correlated with the familiarity model. Regions for which the activation patterns survived the LDC cross-validation are marked in bold

Regions	Cluster Extent	Voxel-level $P$	Pseudo $T$	Coordinates		
				$x$	$y$	$z$
<b>Familiarity</b>						
Left Inferior Frontal Gyrus Opercularis (BA 44)	55	0.001	5.4	-54	5	18
Left Inferior Frontal Gyrus Triangularis (A 45)						
<b>Left Insula</b>	156	0.001	5.34	-30	20	-1
Left Ant Cingulate	106	0.001	4.53	-6	41	14
<b>Left Middle Occipital Gyrus</b>	189	0.001	4.23	-42	-70	11
Right Inferior Frontal Gyrus (BA 47)	360	0.001	4.55	48	38	-9
<b>Right Inferior Frontal Gyrus (BA 44)</b>	360	0.001	6.63	51	11	29
<b>Right Insula</b>	263	0.015	5.47	33	23	6
Right Middle Temporal Gyrus	47	0.020	3.80	63	-19	-9
<b>Right Middle Occipital Gyrus</b>	620	0.001	6.68	39	-70	29

**Author contributions**

**Francesca Carota:** Conceptualization, Investigation, Data curation, Formal Analyses, Methodology, Validation, Visualisation, Interpretation, Writing- Original draft, Review&editing.

**Hamed Nili:** Methodology, Validation. **Friedemann Pulvermüller:** Supervision, Writing.

**Nikolaus Kriegeskorte:** Supervision, Methodology, Writing.

**Data and Code Availability Statement**

The data used in the current study are subjected to restrictions imposed by the administering institution where they were initially collected, the Cognition and Brain Sciences Unit of the MRC and Cambridge University, Cambridge, UK.

For the analyses, code is publicly available <http://www.mrc-cbu.cam.ac.uk/methods-and-resources/toolboxes/license/>