

A Heteromodal Word-Meaning Binding Site in the Visual Word Form Area under Top-Down Frontoparietal Control

Lang Qin,^{1,2*} Bingjiang Lyu,^{3*} Su Shu,^{1,4} Yayan Yin,^{5,6} Xiongfei Wang,^{7,8} Jianqiao Ge,¹ Wai-Ting Siok,² and Jia-Hong Gao^{1,4,9}

¹Center for MRI Research, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China, 100871, ²Department of Linguistics, the University of Hong Kong, Hong Kong, China, 999077, ³Centre for Speech, Language and the Brain, Department of Psychology, University of Cambridge, UK, CB2 3EB, ⁴Beijing City Key Lab for Medical Physics and Engineering, Institute of Heavy Ion Physics, School of Physics, Peking University, Beijing, China, 100871, ⁵Department of Radiology, Xuanwu Hospital, Capital Medical University, Beijing, China, 100053, ⁶Beijing Key Laboratory of Magnetic Resonance Imaging and Brain Informatics, Beijing, China, 100053, ⁷Department of Neurosurgery, Sanbo Brain Hospital, Capital Medical University, Beijing, China, 100093, ⁸Beijing Key Laboratory of Epilepsy, Beijing, China, 100093, and ⁹McGovern Institute for Brain Research, Peking University, Beijing, China, 100871

The integral capacity of human language together with semantic memory drives the linkage of words and their meaning, which theoretically is subject to cognitive control. However, it remains unknown whether, across different language modalities and input/output formats, there is a shared system in the human brain for word-meaning binding and how this system interacts with cognitive control. Here, we conducted a functional magnetic resonance imaging experiment based on a large cohort of subjects (50 females, 50 males) to comprehensively measure the brain responses evoked by semantic processing in spoken and written word comprehension and production tasks (listening, speaking, reading, and writing). We found that heteromodal word input and output tasks involved distributed brain regions within a frontal-parietal-temporal network and focally coactivated the anterior lateral visual word form area (VWFA), which is located in the basal occipitotemporal area. Directed connectivity analysis revealed that the VWFA was invariably under significant top-down modulation of the frontoparietal control network and interacts with regions related to attention and semantic representation. This study reveals that the VWFA is a key site subserving general semantic processes linking words and meaning, challenging the predominant emphasis on this area's specific role in reading or more general visual processes. Our findings also suggest that the dynamics between semantic memory and cognitive control mechanisms during word processing are largely independent of the modalities of input or output.

Key words: controlled semantic cognition; word-meaning binding; visual word form area; word comprehension and production; fMRI

Significance Statement

Binding words and their meaning into a coherent whole during retrieval requires accessing semantic memory and cognitive control, allowing our thoughts to be expressed and comprehended through mind-external tokens in multiple modalities, such as written or spoken forms. However, it is still unknown whether multimodal language comprehension and production share a common word-meaning binding system in human brains and how this system is connected to a cognitive control mechanism. By systematically measuring brain activity evoked by spoken and written verbal input and output tasks tagging word-meaning binding processes, we demonstrate a general word-meaning binding site within the visual word form area (VWFA) and how this site is modulated by the frontal-parietal control network.

Received Oct. 30, 2020; revised Feb. 19, 2021; accepted Feb. 22, 2021.

*L.Q. and B.L. contributed equally to this work.

Author contributions: B.L., J.G. and J.H.G. designed research; L.Q. and B.L. performed research; L.Q., B.L., S.S., Y.Y., X.W. analyzed data; L.Q. wrote first draft of the manuscript; L.Q., B.L., S.S., Y.Y., X.W., J.G., W.T.S., and J.H.G. edited manuscript.

This work was supported by National Natural Scientific Foundation of China Grants 81790650, 81790651, 81727808, 31421003, 81627901, and 31771253; the National Key Research and Development Program of China Grant 2017YFC0108900; the Beijing Municipal Science and Technology Commission

Grant Z181100001518003; and the Capital's Funds for Health Improvement and Research Grant 2020-4-801. We thank the National Center for Protein Sciences at Peking University in Beijing, China, for assistance with data acquisition and analyses.

The authors declare no competing financial interests.

Correspondence should be addressed to Jia-Hong Gao at jjgao@pku.edu.cn.

<https://doi.org/10.1523/JNEUROSCI.2771-20.2021>

Copyright © 2021 the authors

Introduction

Binding words and their meaning as a central language processing is dependent on semantic memory, wherein concepts are extracted from or externalized into forms in different modalities, such as spoken or written words. Semantic memory comprises conceptual associations abstracted from experiences without reference to specific instances, including word meanings, objects and facts (Martin, 2007; Patterson et al., 2007; Binder and Desai, 2011; Forseth et al., 2018). When in the service of language, multiple lead-in processes access the semantic memory repository and drive word-meaning binding to enable verbal comprehension and production (Indefrey and Levelt, 2004; Forseth et al., 2018), which should be modulated by cognitive control (Binder and Desai, 2011). Nevertheless, it is unclear (1) where the modality-invariant word-meaning binding system is located in the brain and (2) how this system is subject to semantic control.

Attempts have been made to identify the neural correlates underlying semantic representational systems across input modalities. A strong resemblance has been demonstrated between two sets of semantic representations recalled by spoken and written stories during language comprehension (Deniz et al., 2019), and the left anterodorsal pars triangularis (Brodmann area 45, BA 45) is significantly correlated with cross-modal semantic similarity encoding (Liuzzi et al., 2017). Furthermore, previous literature suggests that word storage and retrieval mainly rely on ventrolateral temporal lobes, particularly the middle temporal gyrus (MTG), inferior temporal gyrus (ITG), and fusiform gyrus (FG; Binder et al., 2009; Davis and Gaskell, 2009). The ventrolateral temporal lobes include several regions for phonology-semantic-orthography linkage, including the lexical interface (Hickok and Poeppel, 2007; Forseth et al., 2018), the left ventral anterior temporal lobe [ATLv; also referred to as basal temporal language area (BTLA); Binney et al., 2010; Purcell et al., 2014], and the visual word form area (VWFA; Dehaene and Cohen, 2007, 2011). These anatomic and functional areas have long been considered heteromodal integration regions, and recent findings corroborate that posterior parts of the lateral MTG, ITG, and middle FG are centrally involved in representing lexical items independently of modalities (Forseth et al., 2018; Evans et al., 2019; Mattioni et al., 2020). Latest studies with epilepsy patients that used naming tasks to tag the word processing uncovered that the middle occipitotemporal cortex (i.e., FG and ITG) may function as lexical semantic hub and play a crucial part in associating words and their meaning (Forseth et al., 2018; Binder et al., 2020).

Recent research proposed controlled semantic cognition (CSC) framework wherein semantic cognition is dependent on two principal interacting neural systems: one for semantic representation and another for controlling the activity within the representational system according to specific contexts (Lambon Ralph et al., 2017). According to the CSC, concepts arise from multidimensional verbal and nonverbal experiences encoded in modality-specific brain regions and the multimodal hub situated in the bilateral ATLv. The ATLv necessarily mediates the interactions among modality-specific attributes (Patterson et al., 2007; Lambon Ralph et al., 2017) and is engaged across different modalities (verbal and nonverbal semantics, auditory and visual input presentation; Visser et al., 2012; Rice et al., 2015). Prior research on cognitive control in general (O'Reilly et al., 2002; Badre et al., 2005; Badre and D'Esposito, 2009; Cole et al., 2013), semantic retrieval specifically (Thompson-Schill et al., 1997; Wagner et al., 2001), and disordered semantic control (Jefferies

and Lambon Ralph, 2006; Rogers et al., 2015; Thompson et al., 2015) jointly point out that the semantic control network is composed of prefrontal and temporoparietal regions. Recently, a large-scale meta-analysis clearly demonstrated that semantic control is dependent on a left-dominant distributed network consisting of the inferior frontal gyrus (IFG), posterior MTG/ITG, and dorsomedial prefrontal cortex (PFC), with the inferior parietal lobe (IPL) being occasionally activated for its key part in more domain-general cognitive control (Jackson, 2021).

Here, we investigated the existence of a task/modality-invariant word-meaning binding site and illustrated its interaction with semantic control. We hypothesized that semantic hub areas (especially the ventral temporal area) should act as a general word-meaning binding system and activity within this system is modulated by the semantic control network (Fig. 1A).

Materials and Methods

To test our hypotheses, we conducted an fMRI experiment using a paradigm specifically tagging word-meaning association with varying input and output modalities (acoustic or visual; Fig. 1B). We presented concrete nouns referring to familiar objects in daily life as speech sounds, visual words, or pictures to investigate the cortical processing of single-word retrieval.

Experimental design

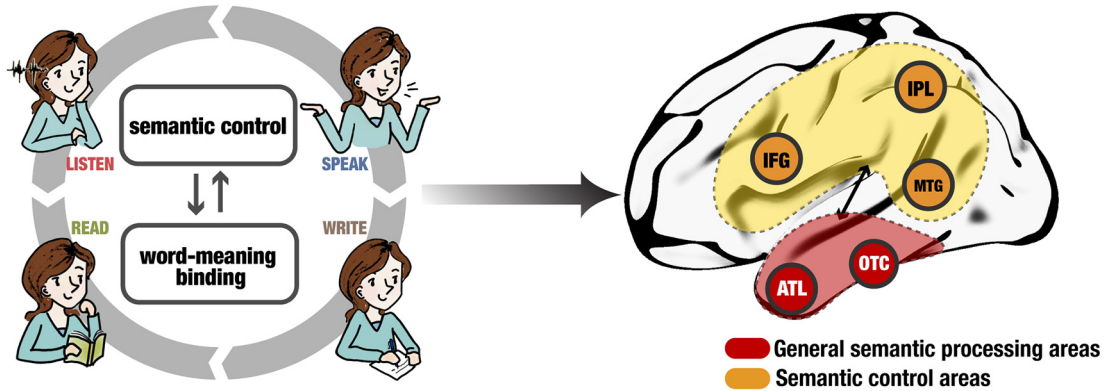
Subjects

Number of subjects ($N = 100$) was determined by a prestudy statistical power analysis with the empirical parameters (intersubject variability = 0.3%, intrasubject variability = 0.75%) and simulation procedure described in a previous study (Desmond and Glover, 2002). The number of time-points in the simulated time series was selected as 128 which corresponded to the shortest one in the four task sections (i.e., reading section, see more details below). The results showed that a sample size of 39 reached a typical 80% statistical power, while a sample size of 100 reached 100% power at the voxel level activation with a signal change of 0.75%. The power provided by a sample size of 100 was still sufficient even with a relatively small signal change (90% power with a signal change of 0.5%). A recent study on fMRI data power prediction also suggested that a sample size of 70 already approached 100% power for the random field theory family-wise error rate inference for peak effects (Durnez et al., 2016). To obtain available data of 100 subjects, 101 healthy native Chinese speakers (50 females; 23.0 ± 2.2 years old) were recruited in this study as paid volunteers. One male subject was excluded because of large head motion (>2.5 mm). None of the subjects reported a history of mental disorder or any kind of language impairment, and they were all right-handed with normal hearing and normal (or corrected) vision. All subjects provided written informed consent approved by the local Institutional Review Board at Peking University.

Task procedure

Our study consists of two fMRI sessions, four sections (i.e., listening, reading, speaking, and writing) in each session, and two tasks (target task and baseline task) in each section. All subjects in this study completed two scan sessions in an MRI scanner on separate days (with an interval of 1.74 ± 1.69 d). Each scan session was composed of four task sections of single-word processing. Each of the four task modalities comprised a target (Lanm, Ranm, Spic, and Wpic) task to evoke explicit semantic processing and a baseline task (Lgdr, Rchr, Ssyl, and Wchr) to control other semantically irrelevant processes such as primary sensory-motor responses or sublexical orthographic encoding. In the listening section, the speech sounds of words or their time-reversed equivalents were presented in the animacy judgment task (Lanm) or gender judgment task (Lgdr), in which subjects were asked to judge whether the word they heard was a creature (for normal speech sounds) or to judge the gender of speakers (for time-reversed speech sounds). Auditory stimuli were presented binaurally using a pair of MRI-compatible headphones, which provided 25-dB/sound pressure level (SPL) attenuation of noise. Subjects were allowed to adjust the sound volume in a short

A Hypotheses



B Experimental paradigm & stimuli

	LISTEN	SPEAK	READ	WRITE
vs.	<p>ann animacy judgement (4blk/ses)</p> <p>gdr gender judgement (4blk/ses)</p>	<p>pic overt picture naming (4blk/ses)</p> <p>syl overt syllable repeating (4blk/ses)</p>	<p>ann animacy judgement (4blk/ses)</p> <p>chr character number judgement (4blk/ses)</p>	<p>pic overt picture naming (4blk/ses)</p> <p>chr overt character copying (4blk/ses)</p>
Semantic	Lanm 	Spic 	Ranm 绵羊	Wpic
Non-semantic	Lgdr 	Ssyl fa so la	Rchr 因特网	Wchr 松

C Experimental procedure

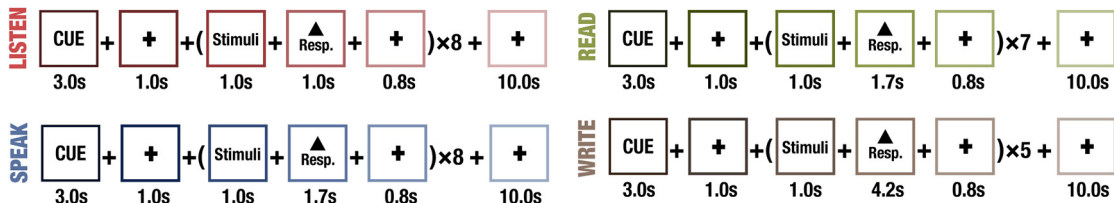


Figure 1. Hypotheses and experimental design. **A**, We hypothesized that regardless of the modalities of word input and output (i.e., spoken or written), word-meaning binding (word retrieval) processes rely on a common general semantic processing system that interacts with a control network (left). The general semantic processing areas (semantic hub) are hypothetically situated within the ATL and/or middle-to-posterior ventral occipitotemporal cortex (vOTC), including areas such as the ITG and FG, under top-down modulation of the control network (right). **B**, Illustration of the experimental paradigm with example stimuli. **C**, Experimental procedure. Each task modality consisted of a target task to evoke semantic responses and a baseline task to control for other semantically irrelevant processes, such as primary sensory-motor responses or sublexical orthographic encoding. Lanm: animacy judgment for normal speech, Lgdr: gender judgment for time-reversed speech, Ranm: animacy judgment for words, Rchr: character number judgment for normal words, Spic: spoken overt picture naming, Ssyl: overt syllable repeating, Wpic: written picture naming, Wchr: normal word copying, resp: response, blk: block, ses: session.

testing scan before formal sessions. The auditory stimuli were presented initially at an 80 dB/SPL. The reading section consisted of an animacy judgment task (Ranm) and character number judgment (Rchr), in which subjects were required to judge whether the picture or word presented referred to a creature or to judge the number of characters in a normal word. All of the judgments were made by pressing corresponding buttons on a handheld keyboard. In the speaking section, pictures and syllable strings were presented in the overt picture naming task (Spic) or overt syllable repeating task (Ssyl), and both names of objects in the pictures and syllable strings were matched in length. When asked to overtly name the picture or repeat the syllable strings, subjects were instructed to speak while keeping their heads as still as possible to avoid head motion. The writing section consisted of an overt picture naming task (Wpic) and a character copying task (Wchr). The subjects were asked to name the pictures by writing down the corresponding words or to copy

normal words or a meaningless symbol. Subjects wrote on an MRI-compatible tablet placed on their abdomen, and they were told to write using only their right hand and keep their other body parts as still as possible. The trajectories of handwriting were recorded in real time. For detailed procedures and the timing of each section, see Figure 1B,C.

Stimuli

Concrete nouns denoting familiar objects in daily life were presented as speech sounds, visual words or pictures to investigate the cortical processing of single-word comprehension and production in different modalities. Chinese words consist of one or more monosyllabic characters, and each square-shaped Chinese character is a combination of radicals or strokes. The speech sounds of nouns were recorded by a male and a female native Chinese speaker in a soundproof studio. The recordings were further edited for length and amplitude, and half of the

recordings were time-reversed to generate acoustic counterparts. There were two sets of auditory stimuli (i.e., one normal set and one time-reversed set) for the two tasks in the listening section. Each set comprised 64 stimuli (half female and half male) that were randomly and evenly assigned to eight blocks. Pictures of objects were chosen from a standardized picture set (Snodgrass and Vanderwart, 1980) of black-and-white line drawings. In the speaking and writing sections, 64 and 40 pictures were presented, respectively, for naming tasks separately. In the reading sections, two sets of 56 normal words were presented separately for the judgment of animacy or number of characters, with words in text form being balanced for word frequency and the number of strokes across tasks. In the speaking section, there were 64 syllable strings for the overt repeating task, and the syllable strings consisted of two or three basic syllables (i.e., do, re, mi, fa, sol, la, ti). In the reading section, each visually presented word was a noun with two or three characters, while single-character nouns were used in the writing section to minimize subjects' movements in the scanner. In the writing section, the copying task involved 40 normal characters.

MRI data acquisition

MRI data were collected using a Siemens Prisma 3T scanner (Siemens Healthineers) with a 20-channel head coil. Thirty-three continuous axial slices that covered the whole brain were acquired using a T_2^* -weighted gradient-echo EPI sequence with the following parameters: TR/TE/FA = 2 s/30 ms/90°, matrix size = 64 × 64, in-plane image resolution = 3.5 × 3.5 mm, slice thickness = 4.2 mm. A high-resolution T_1 -weighted image was acquired for anatomic details with isotropic 1 mm resolution using the MPRAGE sequence (TR/TE/FA = 2.53 s/2.98 ms/7°).

Statistical analysis

fMRI data preprocessing

In short, two fMRI scan sessions were conducted for each subject. There were four sections in each scan session, including listening, speaking, reading, and writing. Each of the four sections included two tasks, one explicit semantic processing task and one baseline task. First-level statistical analysis was conducted for each of the four sections separately with each subject's combined data from the two fMRI scan sessions. For each subject, functional data from the two scan sessions were segmented into four task sections that were separately processed in the following analyses. For each section, realignment, segmentation and normalization were performed using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm>). Functional data were registered to the first volume of the first session through rigid-body realignment and coregistered to T1 images. T1 images were segmented using tissue probability maps, which also returned mutual deformation fields between individual space and Montreal Neurologic Institute (MNI) standard space. Deformation fields were used to normalize functional data to a 2-mm isotropic resolution. A multiple linear regression with nuisance variables was performed on the normalized data. The six head-motion parameters (three for translation and three for rotation) and their first derivatives were included to minimize head-motion effects. The first three principal components of signals from white matter (WM) and CSF and their first derivatives were adopted to regress out nuisance signal components (e.g., cardiac and respiratory effects) based on the CompCor method (Behzadi et al., 2007). Session-related effects were also regressed out in this step. Individual WM and CSF masks were generated by eroding corresponding masks obtained in the segment step (two voxels for WM masks and one voxel for CSF masks). It has been demonstrated that spurious FC arising from motion cannot be totally eliminated by realignment and regression of head-motion parameters (Power et al., 2012). Thus, functional data were further scrubbed by discarding volumes with framewise displacement (FD) larger than 0.5 mm. Isolated volumes with a length less than four volumes after scrubbing were also removed. Data interpolation of the censored data were conducted through a least-squares spectral decomposition of the uncensored volumes; that is, only the "good" data were used to restore signals at censored time points (Power et al., 2014). Only 100 subjects were included in further data analyses because one male subject was excluded because of large head motion (>2.5 mm).

Conjunction analysis

As mentioned before, the target (Lanm, Ranm, Spic, and Wpic) tasks evoked explicit semantic processing, and the baseline tasks (Lgdr, Rchr, Ssyl, and Wchr) controlled other semantically irrelevant processes, such as primary sensory-motor responses or sublexical orthographic encoding (Fig. 1B). Therefore, we first identified the brain activations critically underpinning word-meaning binding processing in four sections (listening, speaking, reading, and writing) by contrasting the target task and baseline task of each section (Tables 2–5). A conjunction analysis (Price and Friston, 1997) was performed at the first level on each task contrast (target task > baseline task) of the four sections using the subject's individual statistical parametric mappings of the minimum t statistic over the contrasts specified in each task. The conjunction analysis allowed us to demonstrate the task-invariant nature of regional responses because it preserved only voxels that were significant (thresholded) in all of the contributing SPM maps to examine responses that were commonly evoked by contrasts of interest. This analysis informed us whether activations were jointly significant in a series of tasks. The results (Tables 6–10) were thresholded at voxel-level $p < 0.001$ with a cluster-level family-wise error (FEW) corrected $p < 0.05$.

Granger causality analysis (GCA)

GCA was used to investigate the causal interactions between regions of interest (ROIs) for each task, and differences between the total causal influence strength flowing into and out of a given ROI were also calculated (i.e., in-degree; Blinowska et al., 2004). Having excluded subcortical areas, cortical areas were selected based on group-level SPM activation maps (target task > baseline task) as ROIs used in GCA. The selection of ROIs encompassed as many ROIs as possible to cover the entire activated regions (Tables 11, 12). ROIs were further classified into large-scale functional networks according to the maximal spatial overlap with Yeo's 7-network cortical parcellation (Yeo et al., 2011). Specifically, we extracted a 10-mm cube for each ROI around its group-level peak coordinates, calculated the proportion of voxels belonging to each network and assigned this ROI to the network with the largest proportion. For individual subjects, the location of each cortical node was defined by searching for the peak activation (uncorrected voxel-level $p < 0.05$) within a 10 mm-radius cube centered at the peak coordinates of group activation results, masked by the corresponding automated anatomic labeling templates (Tzourio-Mazoyer et al., 2002) to avoid contribution from adjacent areas. The time series of each node were extracted as the principal eigenvalue of an activated cluster within a 6-mm radius sphere centered around the individual peak coordinates. Because the specific location of activation may vary across subjects, this procedure guaranteed comparability between models via the application of functional and anatomic constraints in the extraction of time series (Harrison and Tong, 2009). All 100 subjects were included in GCA. For each subject, we extracted time series of eight ROIs for listening, 14 ROIs for speaking, nine ROIs for reading, and 15 ROIs for writing from the explicit semantic processing tasks only (Tables 11, 12). GCA was performed using the Multivariate Granger Causality Toolbox (Barnett and Seth, 2014). The order of the autoregressive model used to obtain the influence measure was determined using the Bayesian information criterion (Schwarz, 1978), and the GC values between pairwise ROIs were calculated. F tests were performed to obtain the statistical significance of inflow/outflow strength at a threshold of $p < 0.05$ with Bonferroni correction for multiple comparisons. We also calculated the sum of GC values flowing into and out of a region and computed the differences between them to measure the information net inflow for a given ROI (in-out degree). A 10,000-times permutation test (the original time series of each ROI was randomly shuffled, and the in-out degree was recalculated) was used to obtain the significance threshold of the in-out degree.

Results

Activation patterns of spoken-/written-language comprehension and production tasks

The behavioral results during the fMRI scan are shown in Table 1. These results indicated that subjects successfully completed all tasks

Table 1. Behavioral results during fMRI scan

Task	Accuracy (%)	Reaction time (ms)	Significant differences
Wchr	99.9 ± 0.6		
Lanm	97.0 ± 3.4	589.8 ± 126.4	rt(Lanm) > rt(Lgdr), $p < 0.001$
Lgdr	97.3 ± 3.8	556.9 ± 143.5	
Ranm	97.4 ± 3.5	589.3 ± 119.7	acc(Ranm) < acc(Rchr), $p = 0.01$ rt(Ranm) > rt(Rchr), $p < 0.001$
Rchr	98.2 ± 3.0	556.1 ± 130.0	
Wpic	96.2 ± 4.0	-	acc(Wpic) < acc(Wchr), $p < 0.001$

The accuracy and reaction time of speaking tasks were not obtained because of technical limitations. Lanm: animacy judgment for normal speech, Lgdr: gender judgment for time-reversed speech, Ranm: animacy judgment for word, Rchr: character number judgment for normal word, Spic: spoken overt picture naming, Ssyl: overt syllable repeating, Wpic: written picture naming, Wchr: normal word copying, rt: reaction time, acc: accuracy.

and that the task-evoked responses were reliable. Contrasted activations tagging the word-meaning binding for spoken and written input and output and their coactivations are shown in Figure 2 and summarized in Tables 2–5. Generally, the activations involved the ventrolateral and posterior medial PFC, posterior dorsal parietal lobe, middle/posterior ventral occipitotemporal regions, and ventromedial occipital lobe (OCC).

Specifically, as shown in Figure 2A, cortical activations (Table 2) for speech comprehension (listening) included the inferior frontal junction (IFJ) and the dorsal premotor cortex (PMd) that falls into BA6, the right anterior middle frontal gyrus (MFGa), the bilateral posterior intraparietal sulcus (IPSp), the precuneus (PCN), and the left middle occipitotemporal sulcus (OTSm). In the spoken-language production task (Fig. 2A; Table 3), more cortical activations were detected than in the comprehension task (i.e., listening). Specifically, apart from regions found for the listening task contrast, other frontal regions were involved, including the bilateral anterior IFG (IFGa), bilateral posterior MFG (MFGp), bilateral presupplementary motor area (preSMA), and left dorsomedial PFC (PFCdm), yet the left PMd, which was found in listening contrast, was absent. For the posterior brain regions found in the listening section, the bilateral PCN and left OTSm were also involved in the speech production task, while bilateral IPSp regions were not. Activations were also found in visual association areas, including the right OTSm and the bilateral OCC.

The written-language comprehension task (reading), as shown in Figure 2A and Table 4, evoked cortical activations overlapping with those found in the spoken-language comprehension and production tasks (listening and speaking). Specifically, similar to the regions found in spoken language tasks, the reading task contrast engaged the bilateral IFJ, the left IFGa, the left PMd, the bilateral preSMA, the left IPSp, the left OTSm and the bilateral OCC. Notably, reading specifically recruited the left sylvian fissure at the end of the parietal temporal border sylvian parietal temporal (SPT), which is suggested to be a sensory-motor integration interface (Hickok and Poeppel, 2007). These activations contain areas previously reported to be involved in reading, namely, Broca's area implicated in language comprehension and production; the ventral occipitotemporal cortex, which includes the VWFA (Cohen et al., 2000; Dehaene et al., 2005; Dehaene and Cohen, 2011), and IPS regions that may be a source of top-down modulation (Planton et al., 2013). Notably, the left IFJ cluster contains part of the left middle frontal region (BA 9/46), which has been identified as a region specific to Chinese reading processing and is correlated with Chinese dyslexia (Siok et al., 2004). For the written-language production (writing) task (Fig. 2A; Table 5), brain area activations also included the aforementioned areas in the other three sections. In terms of bilateral

activations, the writing task in this study activated the bilateral MFGa, preSMA, IFJ, IFGa, IPS, OTSm, and OCC. The left MFGp and SPT were also activated, which conformed with the findings of a previous meta-analysis of writing (Planton et al., 2013). Concretely, cortical activations of our written-language production sections activated writing-specific areas (left PMd, left IFJ, and left IPS), general motor area (preSMA) and areas for linguistic processes (IFGa and OTSm).

To further investigate the common brain substrates shared across comprehension tasks, production tasks, spoken-language tasks, and written-language tasks, conjunction analyses were conducted at the first level on each contrast tagging the association between words and meaning in comprehension tasks (listening, reading) and production tasks (speaking, writing). Comprehension tasks and spoken-language task coactivations were all left hemispheric and involved the left OTSm and the posterior dorsal parietal lobe, while spoken-language tasks additionally recruited the left medial temporal lobe (MTL; Fig. 2B; Tables 6, 8). Furthermore, the production tasks and written-language tasks (Fig. 2B; Tables 7, 9) only coactivated the left OTSm. In addition to activating the left OTSm, the production tasks coactivated the bilateral preSMA, bilateral IPFCdm, bilateral IFGa, bilateral IFJ, bilateral OCC, and left MFGa; moreover, the written-language tasks coactivated the bilateral preSMA, right PFCdm, bilateral IFGa, bilateral IFJ, bilateral IPS, and bilateral OCC.

These findings lend support to the view that the semantic control system overlaps with general executive control and working memory (Lambon Ralph et al., 2017). Bilateral IFJs, especially the right IFJ, are proposed to engage in central cognitive control (i.e., to facilitate goal-directed actions and suppress inappropriate actions; Brass et al., 2005; Braver et al., 2009) and attention control (Zhang et al., 2018; Zhou et al., 2020). Importantly, the IFJ is part of the mid-posterior ventrolateral PFC, which has been repeatedly attested to support domain-general control of semantic selection (Badre et al., 2005). Furthermore, the storage and retrieval of lexical items are suggested mainly to involve the lateral and ventral temporal lobes to which the left OTSm belongs. In addition to the left IFGa, the right IFGa was activated (BA 45/47), and this area has been attested to enhance naming tasks by influencing phonological processing (Naeser et al., 2011). Additionally, the right IFGa is argued to implement response inhibition, a critical aspect of executive functions (Aron et al., 2003). The MFGa, as part of the semantic CON, is also closely related to working memory, enabling previously selected task sets to be held in a pending state for subsequent automatic retrieval and execution on completion of the ongoing task (i.e., cognitive branching; Koechlin and Hyafil, 2007). The posterior parietal parts found here, while being an important part of the human semantic system (Binder et al., 2009; Binder and Desai, 2011), were suggested to support attentional working memory maintenance (Christophel et al., 2018) and episodic memory retrieval (Cavanna and Trimble, 2006).

Spoken-language and written-language comprehension and production tasks coactivate VWFA

To investigate whether there exists a modality-independent word-meaning binding system, a conjunction analysis (Price and Friston, 1997) was conducted to find the region(s) that significantly engaged in the association of words and their meaning across spoken and written language comprehension and production tasks. As shown in Figure 3A, a modality/task overlap significantly coactivated by all four sections (i.e., listening, speaking,

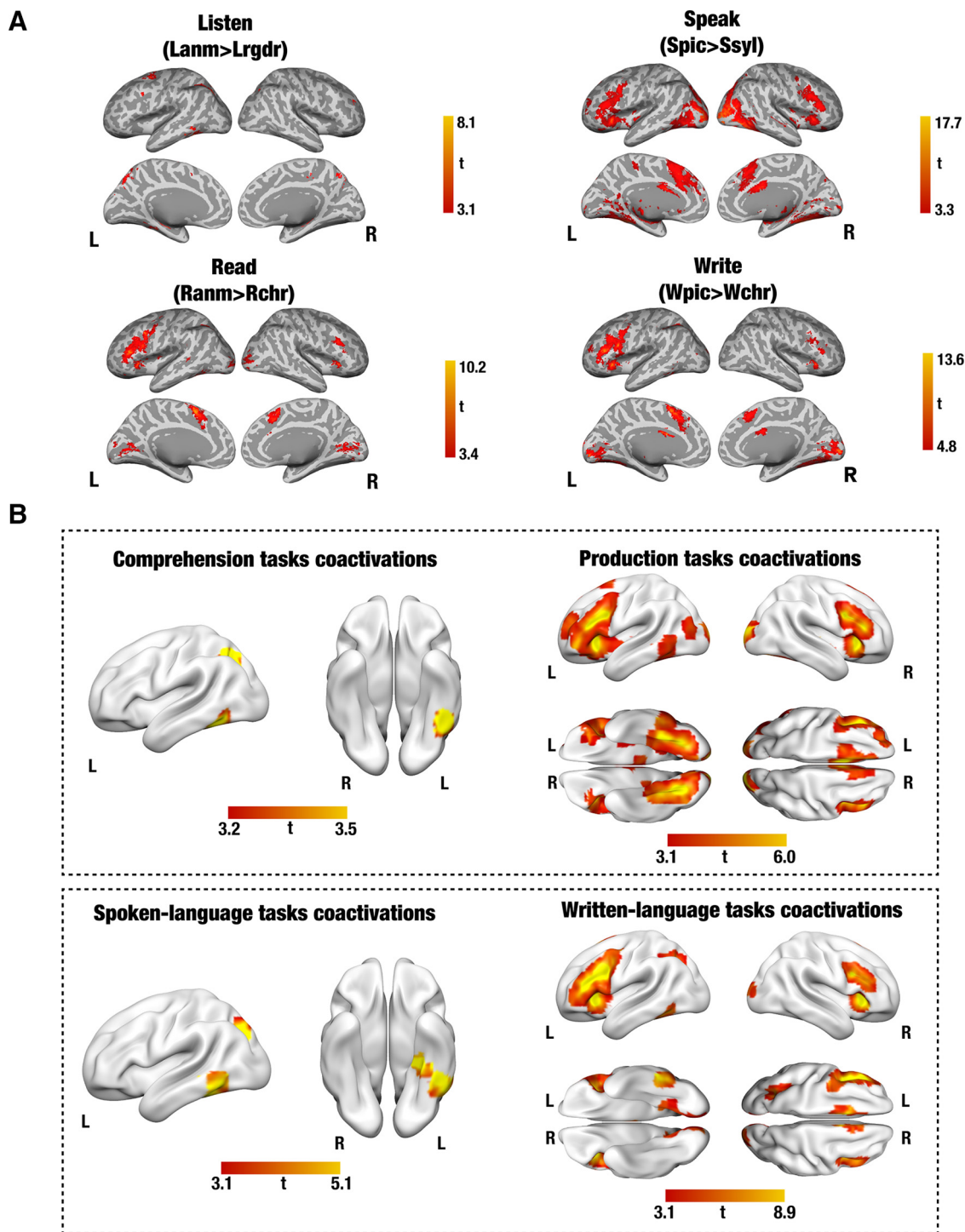


Figure 2. Activations of all four tasks and coactivations revealed by conjunction analysis. **A**, Contrasted brain activations underpinning lexico-semantic processing in spoken-language tasks and written-language tasks. **B**, Coactivations of comprehension and production tasks (top) and coactivations of spoken-language and written-language tasks (bottom). Voxel-level $p < 0.001$ with cluster-level FWE corrected $p < 0.05$.

reading, and writing) was identified in the left OTSm (cluster size: 100 voxels, peak in MNI coordinates: $-46, -46, -14$, thresholded at voxel-level $p < 0.001$ with a cluster-level FWE corrected $p < 0.05$).

This result is consistent with previous findings that lexical item storage and retrieval are primarily pertinent to the ventrolateral temporal regions (Hickok and Poeppel, 2007; Forseth et al., 2018). Intriguingly, the left OTSm overlap fell within the anterior lateral portion of the VWFA (ranging from approximately

Talairach $y = -64$ to -48 or MNI $y = -66$ to -48 , as shown in Fig. 3; Dehaene et al., 2005). The VWFA is a brain site that is selectively attuned to reading and closely related to the visual word form (i.e., the abstract sequence of letters or characters of strokes that composes a written symbol), and it has a pivotal role in the orthography-phonology association (Dehaene et al., 2005; Dehaene and Cohen, 2011). Moreover, the left OTSm overlaps with several conceptual regions that commonly support the linkage of different modality-specific lexical

Table 2. Cortical activations during spoken-language comprehension

Listen (Lanm>Lgdr)	t value	MNI coordinates			AAL	BA
		x	y	z		
R.MFGa	4.43	30	44	26	Frontal_Mid_R	10
L.IFJ	5.36	−40	10	34	Frontal_Mid_L	6/9
	5.34	−28	−4	58	Precentral_L	
	5.07	−26	−4	58	Frontal_Mid_L	
L.PMd	4.70	−24	−2	62	Frontal_Sup_L	6
	4.97	−2	8	50	Supp_Motor_Area_L	
L.IPSp	8.14	−28	−72	36	Occipital_Mid_L	19/39
	5.44	−26	−70	46	Parietal_Sup_L	
	4.63	−30	−52	32	Angular_L	
R.IPSp	4.42	34	−68	40	Occipital_Sup_R	19
L.PCN	5.59	−12	−72	36	PCN_L	7
R.PCN	4.66	12	−76	40	PCN_R	7
	3.61	32	−66	34	Occipital_Mid_R	
L.OTSm	6.55	−50	−48	−14	Temporal_Inf_L	37
	5.78	−56	−46	−6	Temporal_Mid_L	

Cortical activations underlying the word-level processes during spoken-language comprehension were obtained by contrasting the target task (Lanm: animacy judgment for normal speech) and the baseline task (Lgdr: gender judgment for time-reversed speech). The results were thresholded at voxel-level $p < 0.001$ with cluster-level FWE corrected $p < 0.05$. PCN, precuneus; PMd, dorsal premotor cortex; IPSp, posterior intraparietal sulcus; OCC, occipital lobe; OTSm, middle occipitotemporal sulcus; IFGa, anterior inferior frontal gyrus; IFJ, inferior frontal junction; MFGa, anterior middle frontal gyrus; L, left hemispheric; R, right hemispheric; BA, Brodmann's area.

information (Fig. 3; Table 10; Cohen et al., 2002; Hasson et al., 2002; Dehaene et al., 2005; Lacadie et al., 2008; Dehaene and Cohen, 2011; Zhao et al., 2016; Stevens et al., 2017). According to the dual-stream model for speech processing (Hickok and Poeppel, 2007), the left OTSm belongs to the lexical interface area (bilateral ventral temporal regions), which in theory links specific word entries to their lexical/semantic properties (Hickok and Poeppel, 2007), implying that it serves to gather and integrate different sorts of lexical/semantic information. Since lexical knowledge is by no means single modal, the lexical interface must be capable of addressing multimodal information simultaneously to fulfil its role. Additionally, the overlap borders the ventral ATL, which is also referred to as the BTLA (ranging from approximately Talairach $y = -46$ to 17 or MNI $y = -46$ to 20), which is an area that was recently termed the orthography semantic interface region and supports the linkage of modality-specific lexical information types during spoken/written comprehension and production (Purcell et al., 2014).

Previous literature on memory processing and biolinguistics provides indirect evidence that the left OTSm is an ideal site where various types of information converge to be maintained for further processing, with the particular advantage of communicating different language processing routes. The left occipito-temporal area is closely related to working memory maintenance in the absence of perceptual stimulation; it not only displays specificity to multiple categories, especially visual categories, but also retains low-level to high-level abstract representations (Ranganath and D'Esposito, 2001; Harrison and Tong, 2009; Christophel et al., 2017). The temporal association cortex is a cordial interface that communicates with long-term memory for episodic memory retrieval (Vaz et al., 2019), and lesions in this area pertain to certain recognition disorders that are closely related to the functions of polymodal information binding, such as pure alexia (Dehaene et al., 2005). Structurally, the left OTSm is located at the convergence zone of two anatomic pathways for language processing: the dorsal pathway (from meaning to sound) along the arcuate fascicle (AF) and superior longitudinal

Table 3. Cortical activations during spoken-language production

Speak (Spic>Ssyl)	t value	MNI coordinates			AAL	BA
		x	y	z		
L.MFGa	5.55	−30	50	14	Frontal_Mid_L	10
	4.97	−26	48	20	Frontal_Sup_L	
L.MFGp	8.28	−42	4	54	Frontal_Mid_L	6
	7.48	−46	8	50	Precentral_L	
R.MFGp	5.0	44	8	54	Frontal_Mid_R	8
L.IFGa	12.39	−30	28	−4	Frontal_Inf_Orb_L	45/47
	11.75	−28	30	0	Insula_L	
	11.52	−30	30	−2	Frontal_Inf_Tri_L	
	8.50	−30	24	6	Insula_L	
R.IFGa	10.57	34	26	4	Insula_R	45/47
	9.59	32	30	2	Frontal_Inf_Tri_R	
	7.71	34	26	−10	Frontal_Inf_Orb_R	
	4.06	44	20	4	Frontal_Inf_Tri_R	
L.IFJ	12.86	−36	10	28	Frontal_Inf_Oper_L	8/9/46
	9.05	−48	28	20	Frontal_Inf_Tri_L	
	8.28	−42	4	54	Frontal_Mid_L	
	4.92	−42	8	40	Precentral_L	
R.IFJ	8.70	38	8	28	Frontal_Inf_Oper_R	9/46
	7.63	46	12	30	Frontal_Inf_Oper_R	
	5.04	52	24	34	Frontal_Mid_R	
L.PFCdm	7.76	−4	38	44	Frontal_Sup_Medial_L	6/8/9/10
L.preSMA	11.04	−2	16	52	Supp_Motor_Area_L	6/8
R.preSMA	8.73	2	20	52	Supp_Motor_Area_R	6/8/9
L.IPSp	5.96	−20	−72	30	Occipital_Sup_L	7/19
R.IPSp	7.50	28	−72	30	Occipital_Mid_R	7/19/39
	7.20	26	−74	38	Occipital_Sup_R	
L.OTSm	15.04	−30	−46	−20	Fusiform_L	19/22/37
	10.31	−34	−36	−16	Temporal_Inf_L	
	6.99	−46	−76	18	Temporal_Mid_L	
	6.71	−24	−52	−12	Lingual_L	
R.OTSm	13.45	28	−42	−18	Fusiform_R	37
	13.32	50	−64	−10	Temporal_Inf_R	
	10.39	34	−40	−12	ParaHippocampal_R	
L.OCC	13.62	−28	−88	2	Occipital_Mid_L	17/18/19/39
	10.25	−30	−84	−2	Occipital_Inf_L	
	8.18	−10	−100	4	Occipital_Sup_L	
	3.67	−16	−72	6	Calcarine_L	
R.OCC	13.51	36	−82	6	Occipital_Mid_R	17/18/19
	12.93	30	−90	−2	Occipital_Inf_R	
	9.91	24	−90	−2	Lingual_R	
	9.78	20	−96	8	Cuneus_R	

Cortical activations underlying the word-level processes during spoken-language production were obtained by contrasting the target task (Spic: speaking picture naming) and the baseline task (Ssyl: speaking syllable repeating). The results were thresholded at voxel-level $p < 0.001$ with cluster-level FWE corrected $p < 0.05$. PCN, precuneus; IPSp, posterior intraparietal sulcus; IPSa, anterior intraparietal sulcus; OCC, occipital lobe; OTSm, middle occipitotemporal sulcus; IFGa, anterior inferior frontal gyrus; IFJ, inferior frontal junction; preSMA, presupplementary motor area; PFCdm, dorsomedial prefrontal cortex; MFGa, anterior middle frontal gyrus; MFGp, posterior middle frontal gyrus; L, left hemispheric; R, right hemispheric; BA, Brodmann's area.

fascicle (SLF) connecting the temporal lobe and premotor cortices (BA 44, pars opercularis, and PMd) and the ventral pathway (from sound to meaning) along the extreme capsule (EmC) connecting the temporal lobe to the ventrolateral PFC (BA 45/47; Saur et al., 2008). On the one hand, ontogenetic research has demonstrated that the ventral pathway linking the ventrolateral IFG via the EmC to the temporal cortex is detectable in both adults and newborns (Perani et al., 2011). In terms of dorsal pathways, two dorsal pathways are detectable in adult brains: one connects the temporal cortex via the AF/SLF to the posterior portion of the IFG (BA 44), and the other connects the temporal cortex to the PM. However, in the brains of newborns, only the pathway to the PM can be detected. On the other hand,

Table 4. Cortical activations during written-language comprehension

Read (Ranm>Rchr)	t value	MNI coordinates			AAL	BA
		x	y	z		
L.IFGa	7.04	-38	22	-2	Frontal_Inf_Tri_L	45/47
	5.49	-48	18	-6	Frontal_Inf_Orb_L	
L.IFJ	8.09	-46	16	34	Frontal_Inf_Oper_L	8/19/46
	8.00	-44	30	20	Frontal_Inf_Tri_L	
	7.94	-44	10	32	Precentral_L	
	6.47	-50	24	34	Frontal_Mid_L	
R.IFJ	6.57	48	20	28	Frontal_Inf_Tri_R	9/46
	6.05	44	14	28	Frontal_Inf_Oper_R	
	3.94	50	38	20	Frontal_Mid_R	
L.PMd	5.64	-24	-10	50	Frontal_Sup_L	6/9
	5.59	-48	6	48	Precentral_L	
L.preSMA	8.40	-4	12	52	Supp_Motor_Area_L	6/8/32
R.preSMA	8.12	6	14	50	Supp_Motor_Area_R	6/8/32
L.IPSp	5.07	-28	-64	42	Parietal_Inf_L	7/39/19
	4.29	-30	-68	48	Parietal_Sup_L	
	4.24	-28	-56	34	Angular_L	
	3.88	-28	-68	36	Occipital_Mid_L	
L.SPT	5.05	-50	-42	26	SupraMarginal_L	40
L.OTSm	5.78	-46	-52	-18	Temporal_Inf_L	37
	4.81	-48	-40	6	Temporal_Mid_L	
L.OCC	6.58	-28	-90	-4	Occipital_Mid_L	17/18/19
	6.07	-24	-94	-8	Occipital_Inf_L	
	5.90	-16	-72	6	Calcarine_L	
	5.34	-22	-64	2	Lingual_L	
	4.11	-6	-88	14	Cuneus_L	
R.OCC	6.00	34	-84	8	Occipital_Mid_R	18/19
	4.99	32	-90	-2	Occipital_Inf_R	
	5.38	4	-82	4	Calcarine_R	
	5.05	12	-62	8	Lingual_R	

Cortical activations underlying the word-level processes during written-language comprehension were obtained by contrasting the target task (Ranm: animacy judgment for words) and the baseline task (Rchr: character number judgment for normal words). The results were thresholded at voxel-level $p < 0.001$ with cluster-level FWE corrected $p < 0.05$. PCN, precuneus; PMd, dorsal premotor cortex; IPSp, posterior intraparietal sulcus; IPSa, anterior intraparietal sulcus; OCC, occipital lobe; OTSm, middle occipitotemporal sulcus; SPT, sylvian parietal temporal; IFGa, anterior inferior frontal gyrus; preSMA, presupplementary motor area; IFJ, inferior frontal junction; MFGa, anterior middle frontal gyrus; L, left hemispheric; R, right hemispheric.

phylogenetic research has demonstrated that the dorsal pathway with BA 44 as the destination is well developed in human adult brains but ill developed in macaques (Rilling et al., 2008). Thus, since the left OTSm is located within a region that has easy access to multidimensional signals, it is more capable of executing heteromodal integration.

The anterior VWFA is primarily modulated by the frontoparietal semantic control network

Having identified a heteromodal region for general word-meaning association, we examined how it is connected to semantic control across different verbal input and output modalities. We predicted that activity related to the retrieval of word semantics in this region could be modulated by frontoparietal control regions such as the bilateral IFG, an area that has been reported to play a vital part in both semantic control (Noonan et al., 2013; Rogers et al., 2015; Lambon Ralph et al., 2017; Chiou et al., 2018) and general cognitive control (Brass et al., 2005; Stokes et al., 2013).

GCA was applied to investigate the causal interactions between ROIs for each explicit semantic processing task. Briefly, the principle of GCA is based on the predictability of time-varying signals, and one signal is said to Granger cause another if this signal contains information that helps to predict the future behavior of another (Blinowska et al., 2004). Full connections

Table 5. Cortical activations during written-language production

Write (Wpic>Wchr)	t value	MNI coordinates			AAL	BA
		x	y	z		
L.MFGa	6.64	-30	60	16	Frontal_Mid_L	10
R.MFGa	4.94	32	60	18	Frontal_Sup_R	10
	4.24	38	52	24	Frontal_Mid_R	
L.preSMA	13.57	-2	16	52	Supp_Motor_Area_L	6/8/32
R.preSMA	9.06	2	18	56	Supp_Motor_Area_R	6/8/32
L.MFGp	6.61	-46	8	50	Precentral_L	
L.IFJ	12.93	-42	12	28	Frontal_Inf_Oper_L	8/9/46
	10.94	-46	18	28	Frontal_Inf_Tri_L	
	6.75	-48	10	46	Precentral_L	
R.IFJ	8.84	42	12	32	Frontal_Inf_Oper_R	9/46
	8.55	44	32	18	Frontal_Mid_R	
	6.69	48	28	20	Frontal_Inf_Tri_R	
L.IFGa	11.57	-30	30	2	Frontal_Inf_Tri_L	45/47
	10.71	-40	22	-4	Frontal_Inf_Orb_L	
	10.09	-36	24	2	Insula_L	
R.IFGa	7.52	34	24	-8	Frontal_Inf_Orb_R	45/47
	5.73	50	18	-2	Frontal_Inf_Oper_R	
L.IPSp	8.90	-30	-64	42	Parietal_Inf_L	7
	7.90	-28	-64	38	Occipital_Mid_L	
L.SPT	4.11	-58	-46	28	SupraMarginal_L	40
	4.01	-58	-40	34	SupraMarginal_L	
R.IPSa	4.83	36	-54	46	Parietal_Inf_R	40
L.OTSm	10.69	-30	-50	-18	Fusiform_L	37
	7.25	-48	-48	-14	Temporal_Inf_L	
	5.39	-56	-60	2	Temporal_Mid_L	
R.OTSm	9.68	28	-48	-16	Fusiform_R	19/37
	6.36	38	-34	-14	ParaHippocampal_R	
L.OCC	11.07	-12	-98	10	Occipital_Sup_L	17/18/19
	10.96	-10	-84	-12	Lingual_L	
	10.24	-10	-90	-6	Calcarine_L	
	9.93	-24	-72	-10	Fusiform_L	
	5.44	-42	-80	14	Occipital_Mid_L	
	4.09	-54	-62	-10	Temporal_Inf_L	
R.OCC	11.14	14	-90	2	Calcarine_R	17/18/19
	10.74	24	-64	-12	Fusiform_R	
	10.22	8	-76	-10	Lingual_R	
	9.49	16	-96	12	Cuneus_R	
	3.95	46	-78	18	Occipital_Mid_R	

Cortical activations underlying the word-level processes during written-language production were obtained by contrasting the target task (Wpic: written picture naming) and the baseline task (Wchr: normal word copying). The results were thresholded at voxel-level $p < 0.001$ with cluster-level FWE corrected $p < 0.05$. PCN, precuneus; IPSp, posterior intraparietal sulcus; IPSa, anterior intraparietal sulcus; OCC, occipital lobe; OTSm, middle occipitotemporal sulcus; SPT, sylvian parietal temporal; IFGa, anterior inferior frontal gyrus; preSMA, presupplementary motor area; IFJ, inferior frontal junction; MFGa, anterior middle frontal gyrus; MFGp, posterior middle frontal gyrus; L, left hemispheric; R, right hemispheric.

among all ROIs are shown in Figure 4Ai. Figure 4B demonstrates the ROIs that have a direct interaction with the left OTSm. In terms of spoken-language tasks (e.g., listening and speaking), the left OTSm receives causal influence largely from the inferior frontal regions. When performing the listening task, the left IFJ and left posterior inferior parietal region (i.e., IPSp) exert top-down modulation over the left OTSm. For the speaking task, in addition to the left IFJ, the right hemispheric IFJ and ventromedial temporal cortex (i.e., right OTSm) also displayed top-down influence on the left OTSm, with the right OTSm bidirectionally connected to the left OTSm. Similarly, in terms of written-language tasks (e.g., reading and writing), top-down modulation over the left OTSm primarily comes from inferior frontal areas, including the left IFGa and bilateral IFJ, as well as dorsal prefrontal regions and IPL. When performing the reading task, the

Table 6. Coactivations of comprehension tasks

Comprehension tasks	<i>t</i> value	MNI coordinates			AAL	BA
		<i>x</i>	<i>y</i>	<i>z</i>		
L.OTSm	5.48	−46	−46	−14	Temporal_Inf_L	37
L.IPSp	4.98	−26	−66	40	Occipital_Mid_L	39
	3.70	−26	−70	48	Parietal_Sup_L	
	3.15	−32	−58	34	Angular_L	

Conjunction analysis was conducted at the first level on each event of interest in the listening tasks and reading tasks using subjects' individual SPMs of the minimum *t* statistic over the contrasts specified in each task (listening: Lanm>Lgdr, reading: Ranm>Rchr). The results were thresholded at voxel-level $p < 0.001$ with cluster-level FWE corrected $p < 0.05$. Lanm: animacy judgment for normal speech, Lgdr: gender judgment for time-reversed speech, Ranm: animacy judgment for word, Rchr: character number judgment for normal word, BA: Brodman's area.

Table 7. Coactivations of production tasks

Production tasks	<i>t</i> value	MNI coordinates			AAL	BA
		<i>x</i>	<i>y</i>	<i>z</i>		
L.preSMA	11.55	−2	16	52	Supp_Motor_Area_L	6/8/32
R.preSMA	7.49	8	18	48	Supp_Motor_Area_R	6/8/32
L.PFCdm	7.00	0	34	50	Frontal_Sup_Medial_L	6/8
R.PFCdm	7.17	8	22	42	Frontal_Sup_Medial_R	6/8/32
L.IFGa	7.17	−46	20	−8	Frontal_Inf_Orb_L	45/47
R.IFGa	9.94	32	30	0	Frontal_Inf_Tri_R	45/47
	5.60	48	18	−8	Frontal_Inf_Orb_R	
	5.08	50	18	−2	Frontal_Inf_Oper_R	
L.MFGa	5.33	−28	50	14	Frontal_Mid_L	10
L.IFJ	7.25	−36	10	30	Frontal_Inf_Oper_L	46
	8.44	−48	30	20	Frontal_Inf_Tri_L	
R.IFJ	7.78	42	12	30	Frontal_Inf_Oper_R	46
	6.69	44	34	14	Frontal_Inf_Tri_R	
L.OTSm	5.39	−46	−46	−14	Temporal_Inf_L	19/37
	4.51	−52	−52	−4	Temporal_Mid_L	
R.OCC	9.20	26	−68	−12	Fusiform_R	17/19/37/20
	7.40	16	−96	12	Cuneus_R	
	7.24	14	−96	4	Calcarine_R	
	6.00	24	−90	12	Occipital_Sup_R	
L.OCC	8.35	−24	−64	−14	Fusiform_L	17/18/30
	7.39	−10	−98	0	Occipital_Mid_L	
	6.32	−4	−74	−8	Lingual_L	
	3.77	−24	−58	4	PCN_L	
	5.00	6	−82	−4	Lingual_R	

Conjunction analysis was conducted at the first level on each event of interest in the speaking tasks and writing tasks using subjects' individual SPMs of the minimum *t* statistic over the contrasts specified in each task (speaking: Spic>Ssyl; writing: Wpic>Wchr). The results were thresholded at voxel-level $p < 0.001$ with cluster-level FWE corrected $p < 0.05$. Spic: speaking picture naming, Ssyl: speaking syllable repeating, Wpic: written picture naming, Wchr: normal word copying, BA: Brodman's area.

Table 8. Coactivations of spoken-language tasks

Spoken-language tasks	<i>t</i> value	MNI coordinates			AAL	BA
		<i>x</i>	<i>y</i>	<i>z</i>		
L.OTSm	5.12	−48	−46	−14	Temporal_Inf_L	37
L.MTL	4.85	−32	−34	−18	Fusiform_L	36
	4.17	−28	−32	−18	ParaHippocampal_L	
	4.56	−36	−20	−18	Hippocampus_L	
L.IPSp	4.72	−24	−70	36	Occipital_Sup_L	7

Conjunction analysis was conducted at the first level on each event of interest in the listening tasks and speaking tasks using subjects' individual SPMs of the minimum *t* statistic over the contrasts specified in each task (listening: Lanm>Lgdr, speaking: Spic>Ssyl). The results were thresholded at voxel-level $p < 0.001$ with cluster-level FWE corrected $p < 0.05$. Lanm: animacy judgment for normal speech, Lgdr: gender judgment for time-reversed speech, Spic: speaking picture naming, Ssyl: speaking syllable repeating, BA: Brodman's area.

Table 9. Coactivations of written-language tasks

Written-language tasks	<i>t</i> value	MNI coordinates			AAL	BA
		<i>x</i>	<i>y</i>	<i>z</i>		
L.preSMA	8.08	−6	12	50	Supp_Motor_Area_L	6/8/32
R.preSMA	7.11	8	14	50	Supp_Motor_Area_R	6/8/32
R.PFCdm	6.76	8	20	42	Frontal_Sup_Medial_R	
LIFJ	7.60	−44	12	28	Frontal_Inf_Oper_L	8/9/46
	8.18	−46	20	26	Frontal_Inf_Tri_L	
	5.45	−48	10	46	Precentral_L	
R.IFJ	5.87	44	14	28	Frontal_Inf_Oper_R	8/46
	5.76	48	20	28	Frontal_Inf_Tri_R	
	5.64	−48	8	46	Precentral_L	
L.IPSp	5.50	−28	−64	42	Parietal_Inf_L	
	4.42	−30	−58	34	Angular_L	
	3.74	−28	−68	36	Occipital_Mid_L	
L.IPSa	3.70	−32	−54	44	Parietal_Inf_L	39
L.OTSm	5.68	−46	−46	−14	Temporal_Inf_L	37
L.OCC	5.42	−22	−64	4	Calcarine_L	17/30
	5.06	−22	−64	2	Lingual_L	
R.OCC	5.54	14	−68	8	Calcarine_R	17/18/23/30
	5.31	12	−62	8	Lingual_R	
	3.98	20	−94	12	Cuneus_R	
	3.75	24	−90	8	Occipital_Mid_R	

Conjunction analysis was conducted at the first level on each event of interest in the speaking tasks and writing tasks using subjects' individual SPMs of the minimum *t* statistic over the contrasts specified in each task (reading: Ranm>Rchr; writing: Wpic>Wchr). The results were thresholded at voxel-level $p < 0.001$ with cluster-level FWE corrected $p < 0.05$. Ranm: animacy judgment for words, Rchr: character number judgment for normal words, Wpic: written picture naming, Wchr: normal word copying, BA: Brodman's area.

bilateral IFJ and left IFGa exerted significant control over the left OTSm. For the writing task, in addition to the inferior frontal regions found in the reading section, the left preSMA, left MFGa, left MFGp and bilateral IPS were also involved; in contrast with the reading section, the bilateral IFJ was bidirectionally connected to the left OTSm, and the IFGa received input from the left OTSm. The left IPSp and right IPSa sent top-down information to and received bottom-up information from the left OTSm, respectively. The left MFGa also received input from the left OTSm, whereas the left MFGp was bidirectionally connected to the left OTSm. In summary, across four task sections, the left OTSm (i.e., anterior VWFA) primarily received top-down information from prefrontal regions, especially the bilateral IFG. It is worth highlighting that only production tasks (speaking and writing) demonstrated bottom-up modulation from the left OTSm to other regions.

As shown in Figure 4Aii, we further situated the regions involved in explicit semantic processing (i.e., ROIs included in GCA) according to a general parcellation of functional large-scale networks (Yeo et al., 2011), which enabled us to assess these regions in terms of a few domain-general or domain-specific networks, with some of them being consistently found in both resting and task states. With respect to all ROIs and their connections across four sections, production tasks involved more networks, such as default mode network (DMN), as well as more between-network interactions. In terms of ROIs that directly communicated with the left OTSm, the left OTSm, which was classified into the dorsal attention network (DAN), only receives top-down modulation from nodes of the control network (CON) and/or nodes within the DAN itself when performing verbal comprehension tasks (listening and reading). During verbal production tasks (speaking and writing), significant top-down control was also from nodes of CON and/or DAN. However, the speaking section involved bidirectional

Table 10. Peak coordinates of the left OTSm peaks in the current study and the VWFA peak coordinates reported for different writing in other studies

Left OTSm peaks in current study				
Section	Atlas	(x,	y,	z)
1 Cross-section	MNI	(−46,	−46,	−14)
2 Listen	MNI	(−50,	−48,	−14)
3 Speak	MNI	(−42,	−52,	−8)
4 Read	MNI	(−44,	−48,	−12)
5 Write	MNI	(−48,	−48,	−14)
VWFA peaks in other studies				
Writing system	Atlas	(x,	y,	z)
6 Roman alphabet ^a	Talairach	(−43,	−54,	−12)
	MNI*	(−46,	−53,	−20)
7 Simplified Chinese characters ^b	MNI	(−39,	−59,	−12)
8 Roman alphabet ^c	Talairach	(−52,	−49,	−13)
	MNI*	(−57,	−49,	−22)
9 Roman alphabet ^d	Talairach	(−45,	−50,	−12)
	MNI*	(−48,	−49,	−20)
10 Roman alphabet ^e	Talairach	(−42,	−60,	−11)
	MNI*	(−44,	−59,	−19)
11 Simplified Chinese characters ^f	Talairach	(−49,	−53,	−10)
	MNI*	(−52,	−52,	−18)
12 Roman alphabet ^g	Talairach	(−45,	−56,	−16)
	MNI*	(−48,	−55,	−24)
13 Roman alphabet ^h	Talairach	(−44,	−58,	−15)
	MNI*	(−46,	−57,	−23)
14 Japanese kana and kanji ⁱ	Talairach	(−43,	−54,	−12)
	MNI*	(−16,	−53,	−20)

MNI* coordinates are converted from their original Talairach coordinates using Bioimage Suite (Lacadie et al., 2008).

^aVWFA coordinates based on readers of Roman-alphabet languages (Cohen et al., 2002).

^bVWFA coordinates based on readers of simplified Chinese (Zhao et al., 2016).

^cVWFA coordinates based on readers of Roman-alphabet languages (Stevens et al., 2017).

^dVWFA coordinates based on readers of Roman-alphabet languages (Hasson et al., 2002).

^eVWFA coordinates based on readers of Roman-alphabet languages (Hasson et al., 2002).

^{f-h}VWFA coordinates based on readers of Roman-alphabet languages, simplified Chinese, and Japanese kana and kanji (Dehaene and Cohen, 2007).

communication between the left OTSm and visual network (VIS), and the writing section demonstrated bidirectional communication between the left OTSm and nodes of the ventral attention network (VAN) and DMN. Furthermore, the left OTSm has little significant information outflow, whereas this node outputs information to a VIS node (i.e., right OTSm) in speaking and to the bilateral IFJ in reading.

The anterior VWFA has more information input from multiple sources than output in production tasks

Conjunction analyses and GCA analysis jointly indicated that the left OTSm (i.e., anterior VWFA) was primarily modulated by prefrontal regions when critically subserving the binding between words and their meaning regardless of task modality and whether the tasks are receptive or expressive. According to the connectivity pattern shown in Figure 4, it appears that as a word-meaning binding center, in verbal comprehension tasks (listening and reading), the left OTSm appeared to receive all necessary retrieved words and semantic attributes from various code-specific sensory-motor and linguistic resources, whereas in verbal production tasks, the left OTSm seemed to have more bottom-up feedback into other nodes. A possible reason could be that comprehension tasks mainly required mapping between linguistic stimuli to words/concepts, but production tasks triggered mapping from nonlinguistic stimuli to words/concepts, and the words/concepts were mapped to external linguistic forms (speech sounds or written characters), particularly for

Table 11. Coordinates of ROI nodes for GCA analysis (spoken-language tasks)

	t value	MNI coordinates			BA
		x	y	z	
Listen (Lanm>Lgdr)					
L.PCN	5.59	−12	−72	36	7
R.PCN	4.66	12	−76	40	7
L.PMd	4.70	−24	−2	62	6
L.IPSp	8.02	−28	−70	36	39
R.IPSp	4.42	34	−68	40	39
L.OTSm	6.55	−50	−48	−14	37
L.IFJ	5.36	−40	10	34	9
R.MFGa	4.43	30	44	26	19
Speak (Spic>Ssyl)					
L.PCN	5.96	−20	−72	30	7
R.PCN	7.50	28	−72	30	7
R.OTSm	13.45	28	−42	−18	19
L.OCC	11.12	−28	−94	2	18
R.OCC	12.93	30	−90	−2	18
L.OTSm	5.72	−42	−52	−8	37
L.IFGa	14.88	−30	30	−2	47
R.IFGa	10.57	34	26	4	47
L.preSMA	11.04	−2	16	52	8
L.IFJ	12.86	−36	10	28	9
R.IFJ	11.58	38	8	30	9
L.MFGp	8.28	−42	4	54	6
R.MFGp	5.00	44	8	54	6
L.PFCdm	6.80	−2	44	42	6

PCN, precuneus; PMd, dorsal premotor cortex; PFCdm, dorsomedial prefrontal cortex; IPSp, posterior intraparietal sulcus; IPSa, anterior intraparietal sulcus; OCC, occipital lobe; OTSm, middle occipitotemporal sulcus; IFGa, anterior inferior frontal gyrus; IFJ, inferior frontal junction; preSMA, presupplementary motor area; MFGa, anterior middle frontal gyrus; L, left hemispheric; R, right hemispheric; BA, Brodmann's area.

Table 12. Coordinates of ROI nodes for GCA analysis (written-language tasks)

	t value	MNI coordinates			BA
		x	y	z	
Read (Ranm>Rchr)					
L.OCC	6.58	−28	−90	−4	18
R.OCC	6.00	34	−84	8	18
L.PMd	5.64	−24	−10	50	6
L.IPSp	5.07	−28	−64	42	7
L.OTSm	5.67	−44	−48	−12	37
L.SPT	5.05	−50	−42	26	40
L.IFGa	6.61	−48	38	12	46
L.IFJ	7.94	−44	10	32	9
R.IFJ	6.57	48	20	28	46
Write (Wpic>Wchr)					
R.OTSm	9.68	28	−48	−16	19
L.OCC	11.07	−12	−98	10	18
R.OCC	11.14	14	−90	2	18
L.IPSp	8.90	−30	−64	42	7
R.IPSa	4.83	36	−54	46	40
L.OTSm	7.25	−48	−48	−14	37
L.IFGa	11.57	−30	30	2	47
R.IFGa	7.52	34	24	−8	47
L.preSMA	13.57	−2	16	52	8
L.SPT	4.11	−58	−46	28	40
L.MFGa	6.64	−30	60	16	10

PCN, precuneus; IPSp, posterior intraparietal sulcus; IPSa, anterior intraparietal sulcus; OCC, occipital lobe; OTSm, middle occipitotemporal sulcus; SPT, sylvian parietal temporal; IFGa, anterior inferior frontal gyrus; IFJ, inferior frontal junction; MFGa, anterior middle frontal gyrus; MFGp, posterior middle frontal gyrus; preSMA, presupplementary motor area; L, left hemispheric; R, right hemispheric.

the writing naming task in which word form is explicitly required as the output.

To further explore the information input and output of the word-meaning binding site under different tasks and modalities,

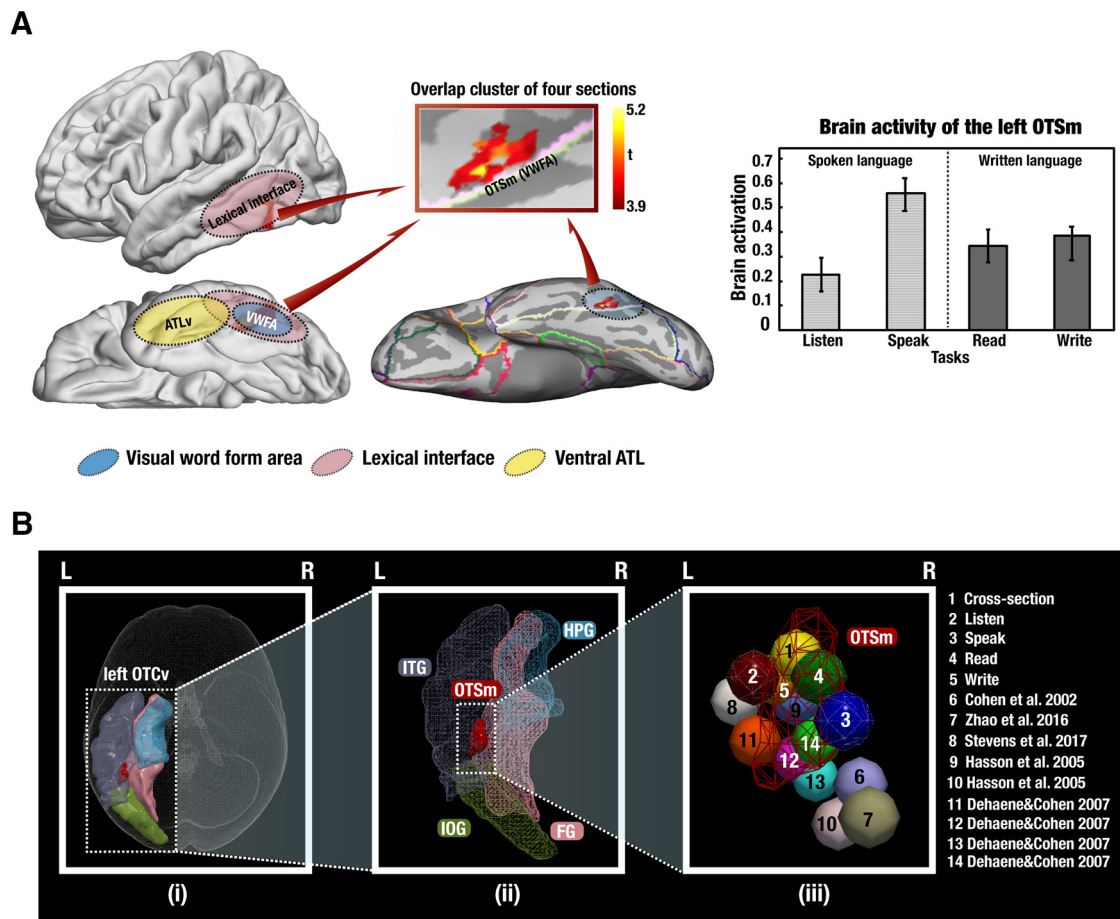


Figure 3. Conjunction analysis revealed that all tasks coactivated VWFA. **A**, The cross-task coactivation that overlapped with VWFA was identified within the left OTSm (voxel-level $p < 0.001$ with cluster-level FWE corrected $p < 0.05$). **Bi**, The left OTCv is the conflux zone of several brain regions. **Bii**, The OTSm is the convergence zone of ITG, IOG, FG and HPG. **Biii**, Coordinates of the cross-section peak (peak 1) and the peaks of contrasts obtained from four sections (peaks 2–5) in the overlap area are close to VWFA peaks reported in previous studies (peaks 6–14); detailed coordinates are shown in Table 10 (the size of the balls has no statistical meaning and was adjusted for demonstration only). OTCv: ventral occipitotemporal cortex; ITG (purple): inferior temporal gyrus; FG (pink): fusiform gyrus; IOG (green): inferior occipital gyrus; HPG (sky blue): hippocampal and parahippocampal gyrus.

we calculated the difference between the sums of GC values flowing into and out of each ROI to measure the information net inflow (i.e., in-out degree). Given its major role assumed during the integration of all necessary retrieved words and semantic attributes, the left OTSm shall have more input than output and a greater net inflow compared with other regions in comprehension tasks, while in production tasks, the outflow from the left OTSm may outweigh its inflow when information of word form is explicitly required in the task output.

The in-out degree results (Fig. 5) revealed that the left OTSm had significant net causal information inflow from multiple sources in listening, speaking, and reading tasks except the writing tasks (permutation test, $p < 0.05$; Bonferroni corrected for multiple comparisons). Interestingly, the left OTSm has much greater input information flow than other ROIs in comprehension tasks (listening and reading), with a smaller difference revealed between the left OTSm and left IPSp for reading. In contrast, a quite different pattern was found for production tasks. For the speaking section, the left OTSm still had more input than output, yet its net inflow no longer remarkably outweighed that in the other regions. For the writing section, the left OTSm showed a reverse pattern with more outflow than inflow. Notably, prefrontal regions assumed to exert semantic control over the left OTSm generally have more output than input (negative or relatively low positive net inflow), such as the bilateral IF and left preSMA.

Discussion

Understanding how the integral capacity of language faculty along with semantic cognition enables the binding between words and their meaning to be prominent in cognitive neuroscience and language sciences. To date, it is unknown whether there is a shared system in the human brain for word-meaning binding and how it interacts with semantic control. To address this issue, we conducted a comprehensive study that simultaneously evaluates word retrieval brain activity across different language modalities and input/output formats. In conclusion, our findings reveal the following. (1) Word-meaning binding under receptive and expressive tasks of different modalities separately involves distributed brain regions within a frontal-parietal-temporal network. (2) The anterior VWFA plays a critical role in word retrieval and arguably is a key site subserving general semantic processes linking words and meaning, which challenges the predominant emphasis on this area's specific role in reading or other general visual processes. (3) The VWFA, as a general word-meaning binding site, receives direct top-down modulation from the frontoparietal CON across all tasks and displays task-specific differences, indicating that the dynamics between the semantic memory and cognitive control mechanism during word processing are largely independent of the modalities of input or output. These results together contribute to a better understanding of how semantic cognition drives word retrieval.

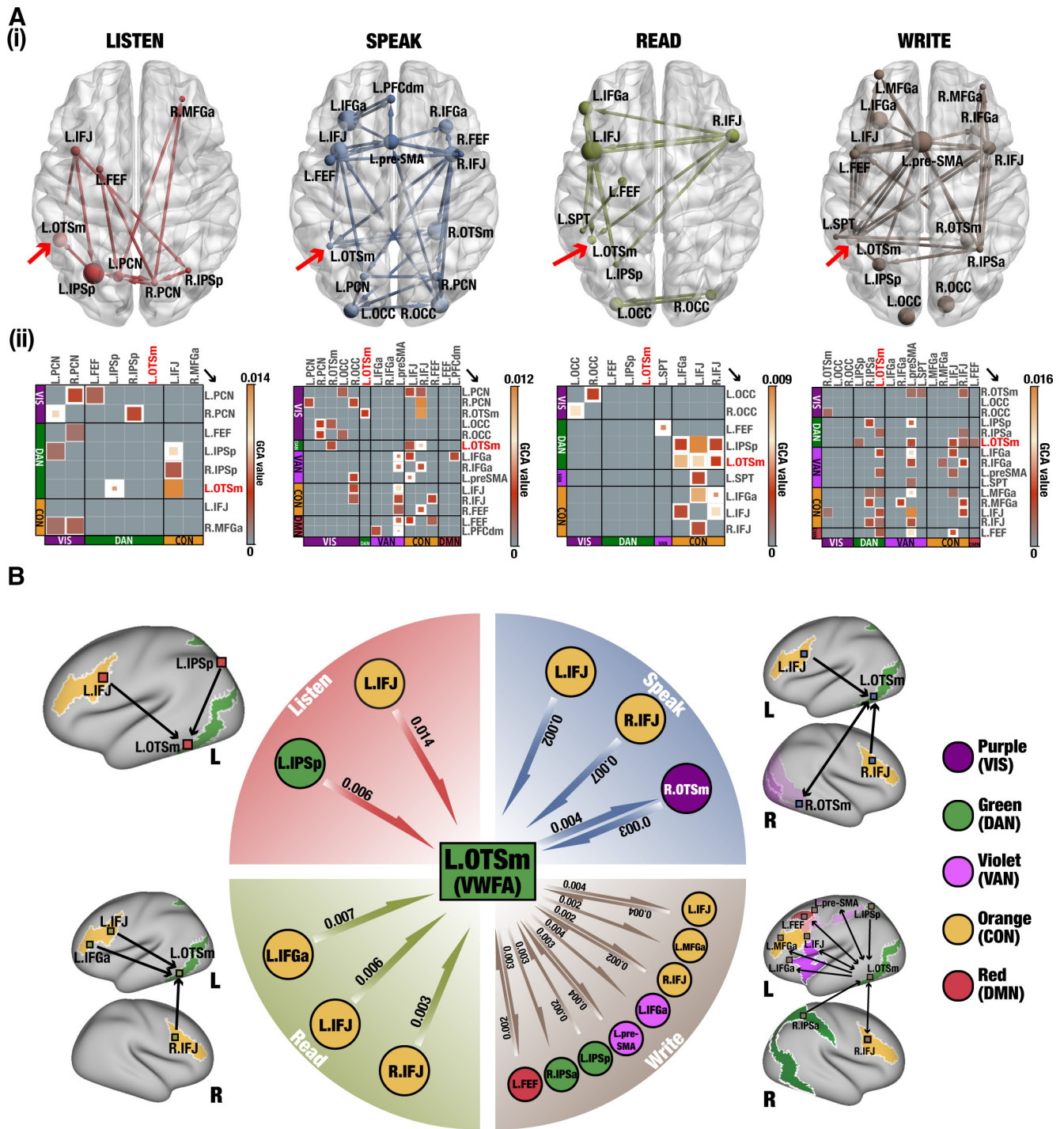


Figure 4. Information flow patterns across four tasks revealed by GC analysis. **Ai**, The causal connectivity graph showing significant directed influences. The size of each node corresponds to its activation (t value). **Aii**, Matrix representations of GCA results for the four tasks. The color of each square represents the GC value of the connection between a certain pair of ROI nodes (FDR corrected $p < 0.05$). **B**, Nodes with significant Granger causal connections to the left OTSm. Circle color indicates large-scale networks defined in Yeo’s 7-network parcellation, including the visual network (VIS), dorsal attention network (DAN), ventral attention network (VAN), frontoparietal control network (CON), and default mode network (DMN). Numbers next to the arrows are GC values that represent the strength of the information flow.

Previous literature has suggested that the left ventral ATL and occipitotemporal regions may serve the multimodal binding involved in word meaning association (Binney et al., 2010; Pulvermüller, 2013; Forseth et al., 2018; Binder et al., 2020). However, our results revealed that the left inferior occipitotemporal region (i.e., the left OTSm or anterior VWFA) acts as a heteromodal word-meaning binding site. For semantic control, a left-lateralized network comprising the IFG, posterior MTG,

posterior ITG and PFCdm has been suggested as its core neural correlate (Jackson, 2021). Accordingly, we uncovered that at the single-word level, the left inferior frontal area, particularly the IFJ, was invariably recruited and modulated the inferior temporal region in a top-down manner, and the right IFJ was also engaged in three sections except the listening section. Importantly, the posterior part of the ITG may also function as a word-meaning association. The left inferior parietal area (i.e., IPSp) was observed

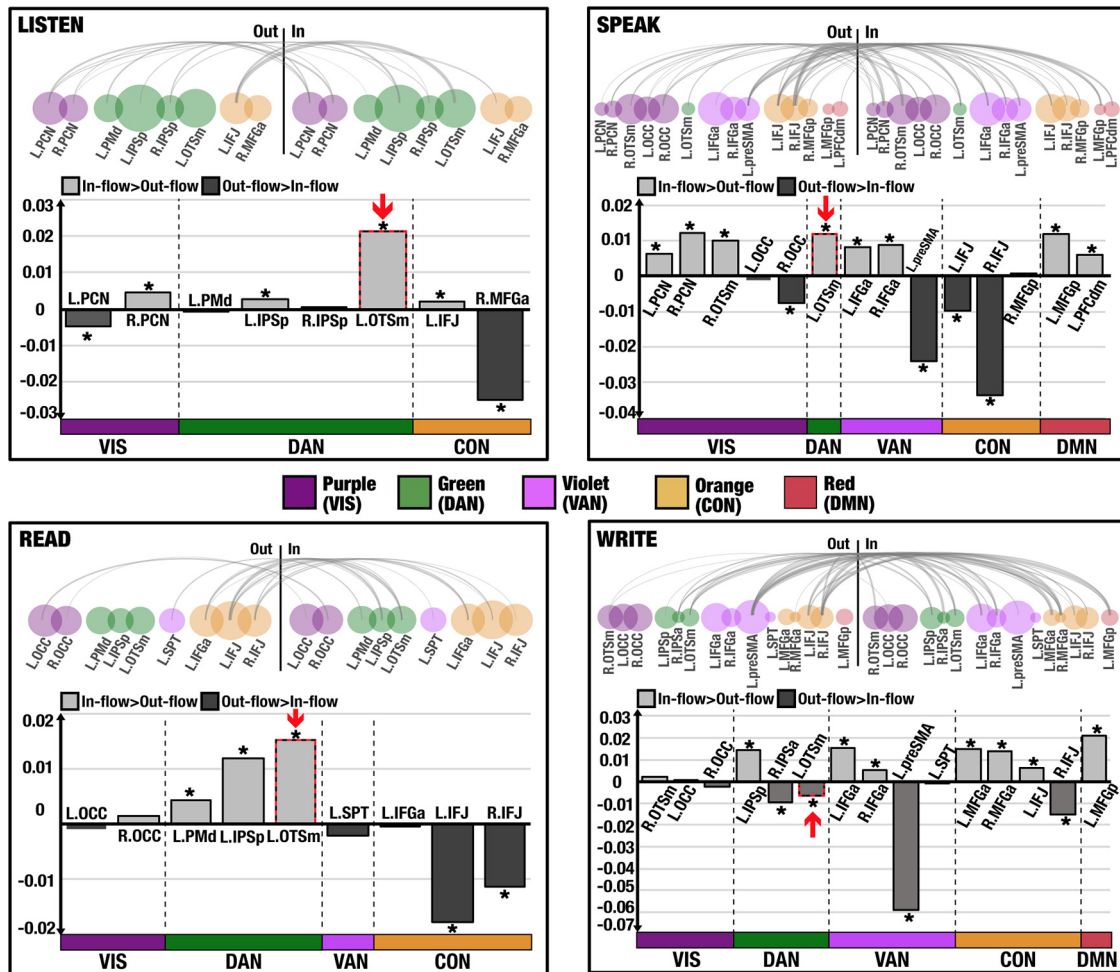


Figure 5. Net Granger causal inflow (in-out degree) of the ROIs of comprehension and production tasks for spoken language and for written language. The asterisks in each diagram denote that the corresponding information in-out degree differs significantly from zero (permutation test, $p < 0.05$; Bonferroni corrected for multiple comparisons). Red arrows designate the left OTSm.

to exert influence over the word-meaning center only in listening and writing, instead of being constantly detected in all four sections. Such occasional involvement is consistent with a priori findings that this region implements more domain-general rather than semantics-specific control and therefore will only be occasionally identified during semantic tasks (Jefferies, 2013; Lambon Ralph et al., 2017; Jackson, 2021). Notably, the posterior MTG was not significantly activated in our study across the four tasks. Past literature has suggested that this area is inclined to boost weakly encoded semantic information retrieval (which requires a greater degree of online exploration of the semantic database) and to become active when control demands within receptive tasks vary (Noonan et al., 2013; Lambon Ralph et al., 2017).

Critically, the current study contributes to solving the VWFA controversy and clarifying its nature. For the particularly strong response of VWFA to written words, there have been two general hypotheses. The first is that VWFA becomes specialized for a single category, subtending visual word formation as prelexical and all visual processing; in line with this assumption, some claimed that this area should be better conceptualized as a region equipped with specific processing characteristics involved in more general visual processes (Price and Devlin, 2003; Gaillard et al., 2006; Glezer et al., 2009; Dehaene and Cohen, 2011; Striem-Amit et al., 2012; Vogel et al., 2014). Another hypothesis is that the VWFA is a special site in the visual cortex with dense interconnections to cortical areas related to language and/or

other important higher level cognitive functions, and it is possible that this hypothesis and the first one are both true (Wandell et al., 2012; Wandell and Le, 2017). Furthermore, prior literature has demonstrated that occipitotemporal brain representations imbued with semantic processing are independent of perceptual modalities (e.g., spoken, written, braille, or sign language; Caramazza et al., 1990; Forseth et al., 2018; Evans et al., 2019; Mattioni et al., 2020). Our results provide strong cross-modality evidence for the view that VWFA is not predominantly or solely dedicated to reading but instead is critical for both spoken-language and written-language processes. More recent findings suggested that the VWFA serves as a link between language and attention (Chen et al., 2019), and the lateral portion of the VWFA may contribute to lexico-semantic access (Bouhali et al., 2019). As the heteromodal word-meaning binding site discovered in this study, the left OTSm (i.e., anterior VWFA) falls in the DAN (Fox et al., 2006), which echoes previous findings that the VWFA is functionally connected to the DAN (Vogel et al., 2012) and is a crucial region for language processing.

Another possible role of the left OTSm, or VWFA, can be identified from the perspective of working memory. By definition, the episodic buffer of working memory congregates and chunks modality-specific representations into a multimodal unity. On the one hand, the buffer temporarily holds relevant information to collectively process these pieces provided that

different aspects and computations of language are suggested to be encoded with different time scales (Sahin et al., 2009; Ding et al., 2016; Sheng et al., 2019). On the other hand, the buffer tackles the gap between various modalities using a multidimensional code, which is required for language processing to gather and integrate information from diverse sources (Rudner and Rönnerberg, 2008). Moreover, the episodic buffer is expected to demonstrate strong modulative interaction primarily with the central executive system but not necessarily under its tight control (Baddeley and Hitch, 2019). While the brain signatures of the episodic buffer per se remain largely unknown, at the core, the buffer for working memory acts as a heteromodal processing center, which is theoretically a role similar to that of the lexical hub for language. Thus, based on cultural recycling theory (Dehaene and Cohen, 2007), we propose that beyond visual word recognition in reading, one of the essential functions for the long-debated VWFA could be cross-modal lexical processing, and VWFA may be recycled from an episodic buffer system by language faculty (Price and Devlin, 2003; Gaillard et al., 2006; Glezer et al., 2009; Dehaene and Cohen, 2011; Striem-Amit et al., 2012; Vogel et al., 2014).

It is also worth noting that while we hypothesized that the ATL_v shall be involved in the linkage of words and their meaning, this critical region is absent. There is accumulated evidence showing that the left ATL is also one of the key regions responsible for combining basic linguistic units into more complex representations (e.g., meaning composition for words in a phrase; Bemis and Pykkänen, 2011; Hagoort, 2019; Pykkänen, 2019). However, in this study, the simple single-word processing tasks we adopted were supposed to have minimal (if any) demands for the basic linguistic featured in the left ATL_v, which may lead to this region being less involved in general. On the other hand, although we had a relatively large sample size that provided reliable statistical power, we cannot rule out the possibility that the absence of ATL_v was to some extent because of signal loss and distortion in ventral temporal regions (Binney et al., 2010; Visser et al., 2012; Lambon Ralph et al., 2017), which needs to be tested with the newly developed techniques to address this issue (Halai et al., 2014; Kundu et al., 2018).

Future research is expected to address important questions pertinent to our findings. The tasks used in this study focused on a single word category (i.e., noun) and were fairly simple (i.e., single-word level processing). To what extent the current findings hold true for other types of words (such as abstract words) or more complicated tasks (e.g., sentence-level processing) is not known. Additionally, the temporal resolution of fMRI limits the tracking of real-time neural dynamics. Particularly, despite a large sample size ($N=100$), the finding of top-down control is based on GCA, which is not an ideal method for fMRI data because of the confounding influence of regional variations in HRF. Thus, it is encouraged to verify these findings using other imaging modalities with a more fine-grained timescale (e.g., milliseconds), such as magnetoencephalography. To tackle these questions, we should more effectively bridge neuroscience and theories from psychology and linguistics to develop appropriate tasks. Additionally, other brain imaging modalities with decent spatiotemporal resolution, such as magnetoencephalography and intracranial electroencephalography, should be adopted.

References

Aron AR, Fletcher PC, Bullmore ET, Sahakian BJ, Robbins TW (2003) Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nat Neurosci* 6:115–116.

Baddeley AD, Hitch GJ (2019) The phonological loop as a buffer store: an update. *Cortex* 112:91–106.

Badre D, D'Esposito M (2009) Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat Rev Neurosci* 10:659–669.

Badre D, Poldrack RA, Paré-Blagoev EJ, Insler RZ, Wagner AD (2005) Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* 47:907–918.

Barnett L, Seth AK (2014) The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *J Neurosci Methods* 223:50–68.

Behzadi Y, Restom K, Liu J, Liu TT (2007) A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37:90–101.

Bemis DK, Pykkänen L (2011) Simple composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *J Neurosci* 31:2801–2814.

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

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

Binder JR, Tong JQ, Pillay SB, Conant LL, Humphries CJ, Raghavan M, Mueller WM, Busch RM, Allen L, Gross WL, Anderson CT, Carlson CE, Lowe MJ, Langfitt JT, Tivarus ME, Drane DL, Loring DW, Jacobs M, Morgan VL, Allendorfer JB, et al. (2020) Temporal lobe regions essential for preserved picture naming after left temporal epilepsy surgery. *Epilepsia* 61:1939–1948.

Binney RJ, Embleton KV, Jefferies E, Parker GJ, Lambon Ralph MA (2010) The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cereb Cortex* 20:2728–2738.

Blinowska KJ, Kuś R, Kamiński M (2004) Granger causality and information flow in multivariate processes. *Phys Rev E Stat Nonlin Soft Matter Phys* 70:050902.

Bouhali F, Bézagu Z, Dehaene S, Cohen L (2019) A mesial-to-lateral dissociation for orthographic processing in the visual cortex. *Proc Natl Acad Sci USA* 116:21936–21946.

Brass M, Derrfuss J, Forstmann B, von Cramon DY (2005) The role of the inferior frontal junction area in cognitive control. *Trends Cogn Sci* 9:314–316.

Braver TS, Paxton JL, Locke HS, Barch DM (2009) Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proc Natl Acad Sci USA* 106:7351–7356.

Caramazza A, Hillis AE, Rapp BC, Romani C (1990) The multiple semantics hypothesis: multiple confusions? *Cogn Neuropsychol* 7:161–189.

Cavanna AE, Trimble MR (2006) The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129:564–583.

Chen L, Wassermann D, Abrams DA, Kochalka J, Gallardo-Diez G, Menon V (2019) The visual word form area (VWFA) is part of both language and attention circuitry. *Nat Commun* 10:5601.

Chiou R, Humphreys GF, Jung J, Ralph MAL (2018) Controlled semantic cognition relies upon dynamic and flexible interactions between the executive 'semantic control' and hub-and-spoke 'semantic representation' systems. *Cortex* 103:100–116.

Christophel TB, Klink PC, Spitzer B, Roelfsema PR, Haynes JD (2017) The distributed nature of working memory. *Trends Cogn Sci* 21:111–124.

Christophel TB, Iamshchinina P, Yan C, Allefeld C, Haynes JD (2018) Cortical specialization for attended versus unattended working memory. *Nat Neurosci* 21:494–496.

Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaux MA, Michel F (2000) The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123:291–307.

Cohen L, Lehéricy S, Chochon F, Lemer C, Rivaud S, Dehaene S (2002) Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain* 125:1054–1069.

Cole MW, Reynolds JR, Power JD, Repovs G, Anticevic A, Braver TS (2013) Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat Neurosci* 16:1348–1355.

- Davis MH, Gaskell MG (2009) A complementary systems account of word learning: neural and behavioural evidence. *Philos Trans R Soc Lond B Biol Sci* 364:3773–3800.
- Dehaene S, Cohen L (2007) Cultural recycling of cortical maps. *Neuron* 56:384–398.
- Dehaene S, Cohen L (2011) The unique role of the visual word form area in reading. *Trends Cogn Sci* 15:254–262.
- Dehaene S, Cohen L, Sigman M, Vinckier F (2005) The neural code for written words: a proposal. *Trends Cogn Sci* 9:335–341.
- Deniz F, Nunez-Elizalde AO, Huth AG, Gallant JL (2019) The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *J Neurosci* 39:7722–7736.
- Desmond J, Glover G (2002) Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. *J Neurosci Methods* 118:115–128.
- Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci* 19:158–164.
- Durnez J, Degryse J, Moerkerke B, Seurinck R, Sochat V, Poldrack R, Nichols T (2016) Power and sample size calculations for fMRI studies based on the prevalence of active peaks. *bioRxiv*
- Evans S, Price CJ, Diedrichsen J, Gutierrez-Sigut E, MacSweeney M (2019) Sign and Speech Share Partially Overlapping Conceptual Representations. *Curr Biol* 29:3739–3747.e5.
- Forseth KJ, Kadipasaoglu CM, Conner CR, Hickok G, Knight RT, Tandon N (2018) A lexical semantic hub for heteromodal naming in middle fusiform gyrus. *Brain* 141:2112–2126.
- Fox MD, Corbetta M, Snyder AZ, Vincent JL, Raichle ME (2006) Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc Natl Acad Sci USA* 103:10046–10051.
- Gaillard R, Naccache L, Pinel P, Clémenceau S, Volle E, Hasboun D, Dupont S, Baulac M, Dehaene S, Adam C, Cohen L (2006) Direct intracranial, fMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron* 50:191–204.
- Glezer LS, Jiang X, Riesenhuber M (2009) Evidence for highly selective neuronal tuning to whole words in the “visual word form area”. *Neuron* 62:199–204.
- Hagoort P (2019) The neurobiology of language beyond single-word processing. *Science* 366:55–58.
- Halai AD, Welbourne SR, Embleton K, Parkes LM (2014) A comparison of dual gradient-echo and spin-echo fMRI of the inferior temporal lobe. *Hum Brain Mapp* 35:4118–4128.
- Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458:632–635.
- Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* 34:479–490.
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Indefrey P, Levelt WJ (2004) The spatial and temporal signatures of word production components. *Cognition* 92:101–144.
- Jackson RL (2021) The neural correlates of semantic control revisited. *Neuroimage* 224:117444.
- Jefferies E (2013) The neural basis of semantic cognition: converging evidence from neuropsychology, neuroimaging and TMS. *Cortex* 49:611–625.
- Jefferies E, Lambon Ralph MA (2006) Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain* 129:2132–2147.
- Koehlin E, Hyafil A (2007) Anterior prefrontal function and the limits of human decision-making. *Science* 318:594–598.
- Kundu S, Kolouri S, Erickson KI, Kramer AF, McAuley E, Rohde GK (2018) Discovery and visualization of structural biomarkers from MRI using transport-based morphometry. *Neuroimage* 167:256–275.
- Lacadie CM, Fulbright RK, Rajeevan N, Constable RT, Papademetris X (2008) More accurate Talairach coordinates for neuroimaging using non-linear registration. *Neuroimage* 42:717–725.
- Lambon Ralph MA, Jefferies E, Patterson K, Rogers TT (2017) The neural and computational bases of semantic cognition. *Nat Rev Neurosci* 18:42–55.
- Liuzzi AG, Bruffaerts R, Peeters R, Adamczuk K, Keuleers E, De Deyne S, Storms G, Dupont P, Vandenberghe R (2017) Cross-modal representation of spoken and written word meaning in left pars triangularis. *Neuroimage* 150:292–307.
- Martin A (2007) The representation of object concepts in the brain. *Annu Rev Psychol* 58:25–45.
- Mattioni S, Rezk M, Battal C, Bottini R, Cuculiza Mendoza KE, Oosterhof NN, Collignon O (2020) Categorical representation from sound and sight in the ventral occipito-temporal cortex of sighted and blind. *Elife* 9:e50732.
- Naeser MA, Martin PI, Theoret H, Kobayashi M, Fregni F, Nicholas M, Tormos JM, Steven MS, Baker EH, Pascual-Leone A (2011) TMS suppression of right pars triangularis, but not pars opercularis, improves naming in aphasia. *Brain Lang* 119:206–213.
- Noonan KA, Jefferies E, Visser M, Lambon Ralph MA (2013) Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *J Cogn Neurosci* 25:1824–1850.
- O’Reilly RC, Noelle DC, Braver TS, Cohen JD (2002) Prefrontal cortex and dynamic categorization tasks: representational organization and neuro-modulatory control. *Cereb Cortex* 12:246–257.
- 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.
- Perani D, Saccuman MC, Scifo P, Anwander A, Anwander A, Spada D, Baldoli C, Poloniato A, Lohmann G, Friederici AD (2011) Neural language networks at birth. *Proc Natl Acad Sci USA* 108:16056–16061.
- Planton S, Jucla M, Roux FE, Démonet JF (2013) The “handwriting brain”: a meta-analysis of neuroimaging studies of motor versus orthographic processes. *Cortex* 49:2772–2787.
- Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59:2142–2154.
- Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE (2014) Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* 84:320–341.
- Price CJ, Friston KJ (1997) Cognitive conjunction: a new approach to brain activation experiments. *Neuroimage* 5:261–270.
- Price CJ, Devlin JT (2003) The myth of the visual word form area. *Neuroimage* 19:473–481.
- Pulvermüller F (2013) How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn Sci* 17:458–470.
- Purcell JJ, Shea J, Rapp B (2014) Beyond the visual word form area: the orthography–semantics interface in spelling and reading. *Cogn Neuropsychol* 31:482–510.
- Pylkkänen L (2019) The neural basis of combinatory syntax and semantics. *Science* 366:62–66.
- Ranganath C, D’Esposito M (2001) Medial temporal lobe activity associated with active maintenance of novel information. *Neuron* 31:865–873.
- Rice GE, Hoffman P, Lambon Ralph MA (2015) Graded specialization within and between the anterior temporal lobes. *Ann NY Acad Sci* 1359:84–97.
- Rilling JK, Glasser MF, Preuss TM, Ma X, Zhao T, Hu X, Behrens TE (2008) The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat Neurosci* 11:426–428.
- Rogers TT, Patterson K, Jefferies E, Ralph MAL (2015) Disorders of representation and control in semantic cognition: effects of familiarity, typicality, and specificity. *Neuropsychologia* 76:220–239.
- Rudner M, Rönnerberg J (2008) The role of the episodic buffer in working memory for language processing. *Cogn Process* 9:19–28.
- Sahin NT, Pinker S, Cash SS, Schomer D, Halgren E (2009) Sequential processing of lexical, grammatical, and phonological information within Broca’s area. *Science* 326:445–449.
- Saur D, Kreher BW, Schnell S, Kümmerer D, Kellmeyer P, Vry M-S, Umarova R, Musso M, Glauche V, Abel S, Huber W, Rijntjes M, Hennig J, Weiller C (2008) Ventral and dorsal pathways for language. *Proc Natl Acad Sci USA* 105:18035–18040.
- Schwarz G (1978) Estimating the dimension of a model. *Ann Statist* 6:461–464.
- Sheng J, Zheng L, Lyu B, Cen Z, Qin L, Tan LH, Huang M-X, Ding N, Gao J-H (2019) The cortical maps of hierarchical linguistic structures during speech perception. *Cereb Cortex* 29:3232–3240.

- Siok WT, Perfetti CA, Jin Z, Tan LH (2004) Biological abnormality of impaired reading is constrained by culture. *Nature* 431:71–76.
- Snodgrass JG, Vanderwart M (1980) A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J Exp Psychol Learn* 6:174–215.
- Stevens WD, Kravitz DJ, Peng CS, Tessler MH, Martin A (2017) Privileged functional connectivity between the visual word form area and the language system. *J Neurosci* 37:5288–5297.
- 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.
- Striem-Amit E, Cohen L, Dehaene S, Amedi A (2012) Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron* 76:640–652.
- Thompson HE, Robson H, Lambon Ralph MA, Jefferies E (2015) Varieties of semantic ‘access’ deficit in Wernicke’s aphasia and semantic aphasia. *Brain* 138:3776–3792.
- 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.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15:273–289.
- Vaz AP, Inati SK, Brunel N, Zaghoul KA (2019) Coupled ripple oscillations between the medial temporal lobe and neocortex retrieve human memory. *Science* 363:975–978.
- Visser M, Jefferies E, Embleton KV, Lambon Ralph MA (2012) Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *J Cogn Neurosci* 24:1766–1778.
- Vogel AC, Miezin FM, Petersen SE, Schlaggar BL (2012) The putative visual word form area is functionally connected to the dorsal attention network. *Cereb Cortex* 22:537–549.
- Vogel AC, Petersen SE, Schlaggar BL (2014) The VWFA: it’s not just for words anymore. *Front Hum Neurosci* 8:88.
- Wagner AD, Paré-Blagoev EJ, Clark J, Poldrack RA (2001) Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31:329–338.
- Wandell BA, Le RK (2017) Diagnosing the neural circuitry of reading. *Neuron* 96:298–311.
- Wandell BA, Rauschecker AM, Yeatman JD (2012) Learning to see words. *Annu Rev Psychol* 63:31–53.
- Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, Fischl B, Liu H, Buckner RL (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* 106:1125–1165.
- Zhang X, Mlynaryk N, Ahmed S, Japee S, Ungerleider LG (2018) The role of inferior frontal junction in controlling the spatially global effect of feature-based attention in human visual areas. *PLoS Biol* 16:e2005399.
- Zhao L, Chen C, Shao L, Wang Y, Xiao X, Chen C, Yang J, Zevin J, Xue G (2016) Orthographic and phonological representations in the fusiform cortex. *Cereb Cortex* 27:5197–5210.
- Zhou L, Zhen Z, Liu J, Zhou K (2020) Brain Structure and Functional Connectivity Associated with Individual Differences in the Attentional Blink. *Cerebral Cortex* 30:6224–6237.