A common neural code for meaning in discourse production and comprehension

Tanvi PATEL

Matías MORALES

Martin J. PICKERING

Paul HOFFMAN*

School of Philosophy, Psychology & Language Sciences, University of Edinburgh, UK

* Correspondence to:

Dr. Paul Hoffman

School of Philosophy, Psychology & Language Sciences,

University of Edinburgh,

7 George Square, Edinburgh, EH8 9JZ, UK

Tel: +44 (0) 131 650 4654

Email: p.hoffman@ed.ac.uk

Acknowledgements

PH was supported by a BBSRC grant (BB/T004444/1). Imaging was carried out at the Edinburgh Imaging Facility (www.ed.ac.uk/edinburgh-imaging), University of Edinburgh, which is part of the SINAPSE collaboration (www.sinapse.ac.uk). For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Abstract

How does the brain code the meanings conveyed by language? Neuroimaging studies have investigated this by linking neural activity patterns during discourse comprehension to semantic models of language content. Here, we applied this approach to the production of discourse for the first time. Participants underwent fMRI while producing and listening to discourse on a range of topics. We quantified the semantic similarity of different speech passages and identified where similarity in neural activity was predicted by semantic similarity. A widely distributed and bilateral network, similar to that found for comprehension, showed such effects when participants produced their own discourse. Critically, cross-task neural similarities between comprehension and production were also predicted by similarities in semantic content. These results indicate that discourse semantics engages a common neural code during both comprehension and production. Furthermore, common coding in right-hemisphere regions challenges the idea that language production processes are strongly left-lateralised.

Introduction

Both the production and comprehension of language are rooted in our conceptual knowledge of the world. At the neural level, this semantic processing is carried out through the activation of a widely distributed network of systems that store and retrieve knowledge [1-3]. Traditionally, work on semantic representation has linked broad semantic categories (e.g., animals vs. tools) to regional engagement of semantic processing areas [4, 5]. But more recently, the development of sophisticated multivariate pattern analysis (MVPA) techniques has allowed researchers to examine how conceptual knowledge is encoded across distributed patterns of neural activation [6, 7]. These methods have moved the field from a coarse category-based approach to semantic representation into a fine-grained multidimensional understanding of the neural coding of meaning. They have also expanded the space of investigation from individual words and objects into the rich array of concepts and situations that occur in natural discourse. In this study we use Representational Similarity Analysis (RSA), an MVPA technique that tests how similarity in the properties of stimuli can predict similarity in the neural activation patterns they elicit [7, 8]. Previous studies have used RSA at the single-word level to investigate how semantic content is coded in the brain [9-12]. The present study applies this approach to the discourse level and asks a critical but under-investigated question: is the coding of semantic content similar when people produce their own narratives, compared to when they listen to another person's speech? In other words, does meaning in language comprehension and production rely on a common neural code?

MVPA has been used to investigate questions of semantic representation at different levels of stimulus complexity. Studies using stimuli at the single word [e.g., 10, 11, 13] and sentence level [e.g., 14, 15, 16] have demonstrated that patterns of neural activation can be predicted from vector-based semantic models. These include models that code conceptual relationships based on associated sensory-motor experiences [17, 18] and models of natural language processing, which derive semantic representations from lexical statistics in large text corpora [19-21]. Beyond the word and sentence levels, studies have used voxel-wise [22, 23] and multivariate approaches to examine language comprehension at the discourse level. Such

studies use more naturalistic stimuli, such as written stories [24, 25] spoken narratives [22, 26, 27] or movies [28]. These studies have revealed a widely distributed set of frontal, temporal, and parietal areas in which activation patterns can be predicted by the semantic content of language.

The critical question in our study is the degree to which these neural codes are similar across different language tasks – specifically, comprehending versus producing speech. MVPA studies have shown that the neural representations of concepts are robust to changes in the stimuli used to evoke them. In other words, concepts elicit similar patterns of activation in the brain's semantic network, irrespective of the stimuli used to evoke them. This cross-stimulus generalisation has been observed for object names vs. object pictures [11, 29]; for written vs. auditory words [30]; for stories vs. movie stimuli [28]; and for event descriptions that vary in lexical and syntactic content [31]. This is true even across languages – neural representations of concepts show cross-linguistic invariance at the individual word [32] and story-level [25]. These findings are consistent with contemporary accounts of semantic cognition, which hold that higher-order conceptual knowledge is supported by "hub" regions that permit generalisation of knowledge across contexts, exemplars and tasks [1, 2, 33, 34].

In the present study, we test whether the same kind of generalisation occurs when comparing activation evoked during different language tasks. Theories of language production and comprehension suggest that these are interwoven processes that recruit overlapping neural networks [35-38]. At the level of discourse, both comprehension and production are thought to involve the construction of situation models which encode the entities and events described [35, 39]. In addition, one particular view postulates that comprehension uses a prediction-by-production mechanism, whereby the listener covertly imitates the linguistic form of the speaker's utterances in order to derive their intentions, and then runs this through their production system to predict the speaker's upcoming utterances [38, 39]. However, there is currently limited evidence that the neural representations engaged when people produce discourse are similar to those elicited when they listen to it. This is in part because, while MVPA studies of discourse comprehension are now commonplace, analogous studies of discourse production are rare.

Univariate neuroimaging studies suggest some differences in the neural networks activated during comprehension and production. While production and comprehension of speech activate common language areas in the left hemisphere, narrative comprehension tends to activate right frontal and temporal regions more than production [40-42]. These findings align with theoretical views that while speech comprehension engages a bilateral network, language production is more left-lateralised [43-47]. More recently, a series of studies have investigated correlations in neural activity between a speaker producing a story and a listener comprehending it [42, 48-50]. These indicate a more bilateral picture, with significant production-comprehension temporal coupling in left and right prefrontal, temporal and inferior parietal cortices. This inter-subject coupling suggests a large degree of shared neural processing across comprehension and production. However, these studies have investigated correlations between different individuals and not within the same individual performing different language tasks. In addition, inter-subject correlations in neural activity could be due to shared processes at any level of the language system and do not indicate where commonalities are due specifically to its semantic content.

Here, we used RSA to investigate the degree to which neural patterns during speech production and comprehension align with a single model of semantic content. In the scanner, participants listened to pre-recorded speech samples, and produced extended passages of speech on a variety of topics. We used a distributional semantic model [latent semantic analysis; 21] to quantify the degree to which different speech chunks contained similar semantic content. We then used RSA to assess how well semantic similarity could predict similarity in the neural responses evoked during different periods of heard or spoken language. Critically, as well as performing these analyses separately for comprehension and production data, we conducted a cross-task analysis that looked at similarities between comprehension and production periods (within participants). In this way, we were able to identify brain regions which share a common neural code for semantic content during speaking and listening.

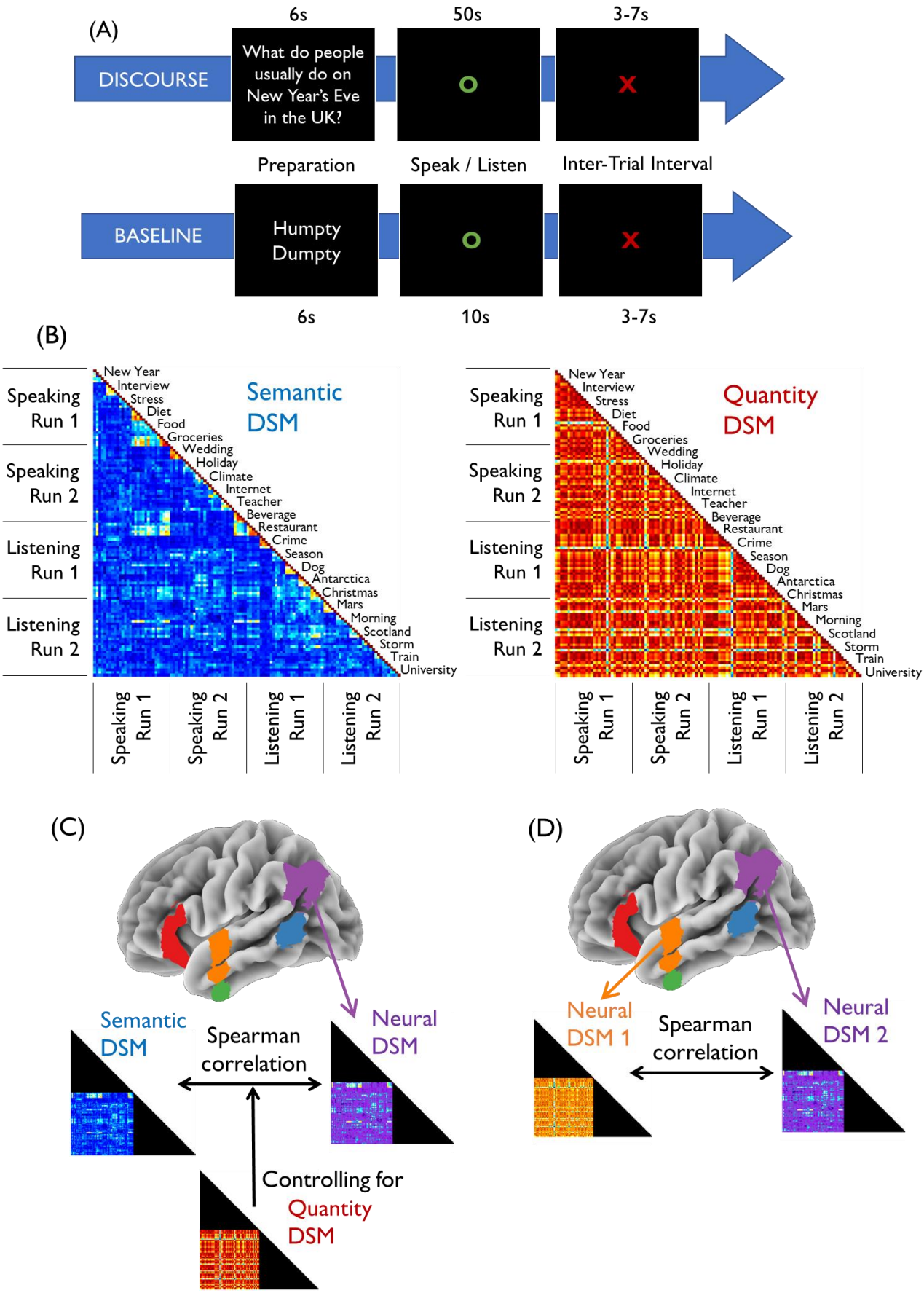
We used a searchlight approach to seek semantic-coding regions across the whole brain. In addition, we investigated effects in a set of regions of interest known to be involved in semantic processing. These comprised parts of the anterior temporal lobes, which act as a hub for representation of semantic knowledge [2, 33], inferior prefrontal and posterior

temporal regions that regulate access to this knowledge [51, 52] and the angular gyrus, a key node in the default mode network which is involved in constructing mental models of events and ongoing experiences [1, 53]. By investigating effects in the left and right hemisphere homologues of these regions, we were also able to probe the extent of lateralization during production and comprehension of speech.

Results

We performed a series of analyses on fMRI data in which participants produced and listened to speech samples on a range of topics (see Figure 1). After reporting basic characteristics of the language samples, we report the results of univariate analyses which investigated the degree to which brain regions were engaged during discourse comprehension and production, compared with a baseline of automatic speech. Following this, we use RSA to identify regions in which similarities in activation patterns are predicted by similarities in the semantic content of language. As well as performing these analyses separately on listening and speaking data, we perform a cross-task analysis that tests for semantic neural coding that generalises across language tasks. Finally, we directly compare neural similarity patterns across semantic-related brain regions to investigate the degree to which different regions are similarly influenced by discourse content.

Figure 1. (A) Structure of trials in the experiment. On each discourse trial, participants were presented with a topic prompt for 6s and were then required to either speak about this topic for 50s or listen to a recording of another person speaking about it for 50s. In the baseline condition, participants either recited or listened to a well-known British nursery rhyme. (B) Example semantic and quantity dissimilarity matrices. Topic labels indicate different discourse topics, each of which was divided into five 10s segments. The semantic DSM codes how similar speech segments are in lexical-semantic content. The quantity DSM (used as a control) codes how similar segments are terms of the number of words they contain. (C) Procedure for main RSA analyses. The analysis tested the relationship between a neural DSM, generated by comparing local activation patterns for different speech segments, and the semantic DSM, while controlling for the effects of the quantity DSM. (D) Procedure for comparison of neural DSMs. This analysis test the relationship between two neural DSMs obtained from different brain regions. Speak1 = speech production run 1; Speak2 = speech production run 2; Listen1 = comprehension run 1; Listen2 = comprehension run 2. DSM = dis-similiarity matrix.


(A)
6s
50s
3-7s
DISCOURSE
What do people usually do on New Year's Eve in the UK?
Preparation
Speak / Listen
Inter-Trial Interval
BASELINE
Humpty Dumpty
6s
10s
3-7s
(B)
Speaking Run 1
Speaking Run 2
Listening Run 1
Listening Run 2
Semantic DSM
Quantity DSM
New Year
Interview
Stress
Diet
Food
Groceries
Wedding
Holiday
Climate
Internet
Teacher
Beverage
Restaurant
Crime
Season
Dog
Antarctica
Christmas
Mars
Morning
Scotland
Storm
Train
University
(C)
Semantic DSM
Spearman correlation
Neural DSM
Controlling for Quantity DSM
(D)
Neural DSM 1
Spearman correlation
Neural DSM 2

Characteristics of speech: In the discourse production task, participants spoke about a series of topics for 50s at a time (see Method for details). They produced a mean of 129 words per topic (SD = 37, range = 67-320). In the discourse comprehension task, participants listened to recordings of another person speaking about a different set of topics. These recordings contained a mean of 141 words per topic (SD = 20, range = 102-170). Example speech samples from comprehension and production tasks are provided in Supplementary Materials .After scanning, participants received 12 comprehension questions relating to the speech presented in the listening task. They answered 10/12 questions correctly on average. They also provided audibility rating with a mean rating of 5.5/7, suggesting that they were able to understand the discourse samples presented in the scanner.

Univariate analyses: Whole-brain activation maps are shown in Figure 2. These show effects for each discourse task relative to a "low-level" speech baseline (reciting/listening to a familiar nursery rhyme). The baseline conditions involved language processing and so had similar perceptual and motor demands to the discourse tasks, but they did not require participants to process novel, meaningful verbal information. In line with previous studies, the results indicate close correspondence in the areas recruited for production and comprehension of novel discourse, particularly in the left hemisphere. Both tasks activated similar left-lateralised networks associated with semantic processing, including lateral and medial prefrontal cortex, lateral temporal cortex, the ventral anterior temporal lobe and the angular gyrus. Listening produced significantly greater activation than speaking in right prefrontal regions, left superior temporal sulcus and bilateral post-central gyrus. Speech production was associated with greater activity in the cerebellum, medial prefrontal cortex and the occipital lobe. The overall picture, however, was that participants engaged broadly similar networks whether they were speaking or listening.

We also assessed activation in five regions of interest (ROIs) that are frequently implicated in semantic processing, and their right-hemisphere homologues (see Figure 3A). Results are shown in Figure 3B and reveal a left-lateralised pattern for both tasks. 2 x 5 x 2 ANOVA showed a main effect of hemisphere ($F(1,24) = 64.4$, $p < .001$; left > right) as well as ROI ($F(4,96) = 20.7$, $p < .001$), and task ($F(1,24) = 5.7$, $p = .026$; listening > speaking). All of

the interactions between these factors were also significant ($F > 6.9$, $p < .001$). FDR-corrected post-hoc tests indicated that activation was left-lateralised for both tasks in most regions (see Figure 3B). In the IFG, pMTG and AG, this effect was larger in the speaking task, with right-hemisphere regions deactivating during production, relative to the baseline speech condition. Similar results were obtained when discourse was contrasted against rest rather than baseline speech conditions (see Supplementary Figure 1). Thus, univariate analyses suggest that the left hemisphere is dominant in semantic aspects of discourse processing, particularly when participants produced, rather than heard, discourse.

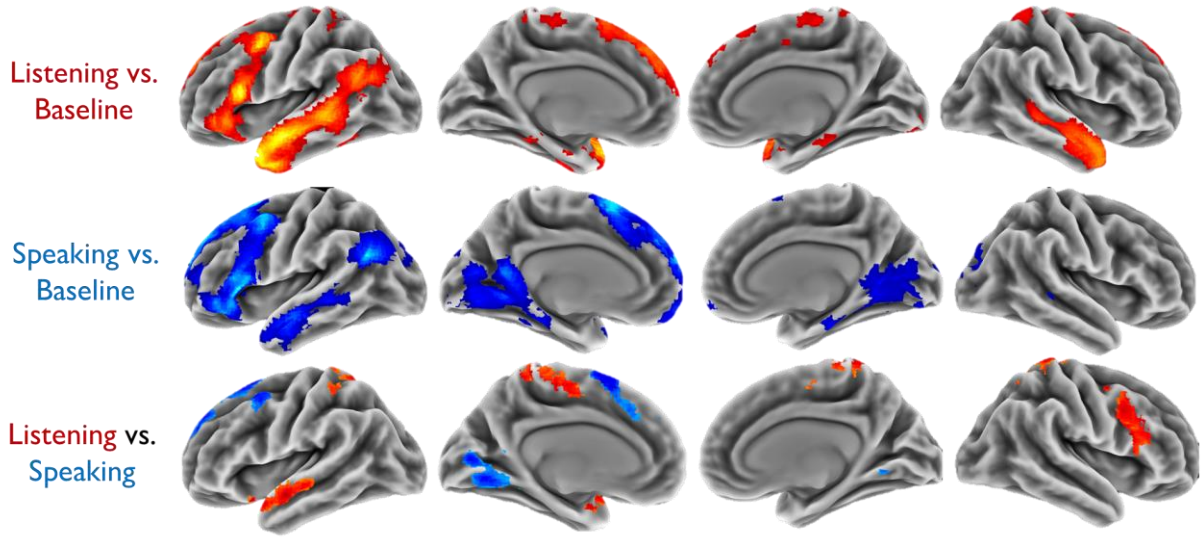


Figure 2: Univariate activation contrasts. Areas activated by each discourse task compared with its baseline, and for the direct contrast of speaking vs. listening. Results are shown at cluster-corrected $p < 0.05$. Peak activation co-ordinates are reported in Supplementary Table 1.

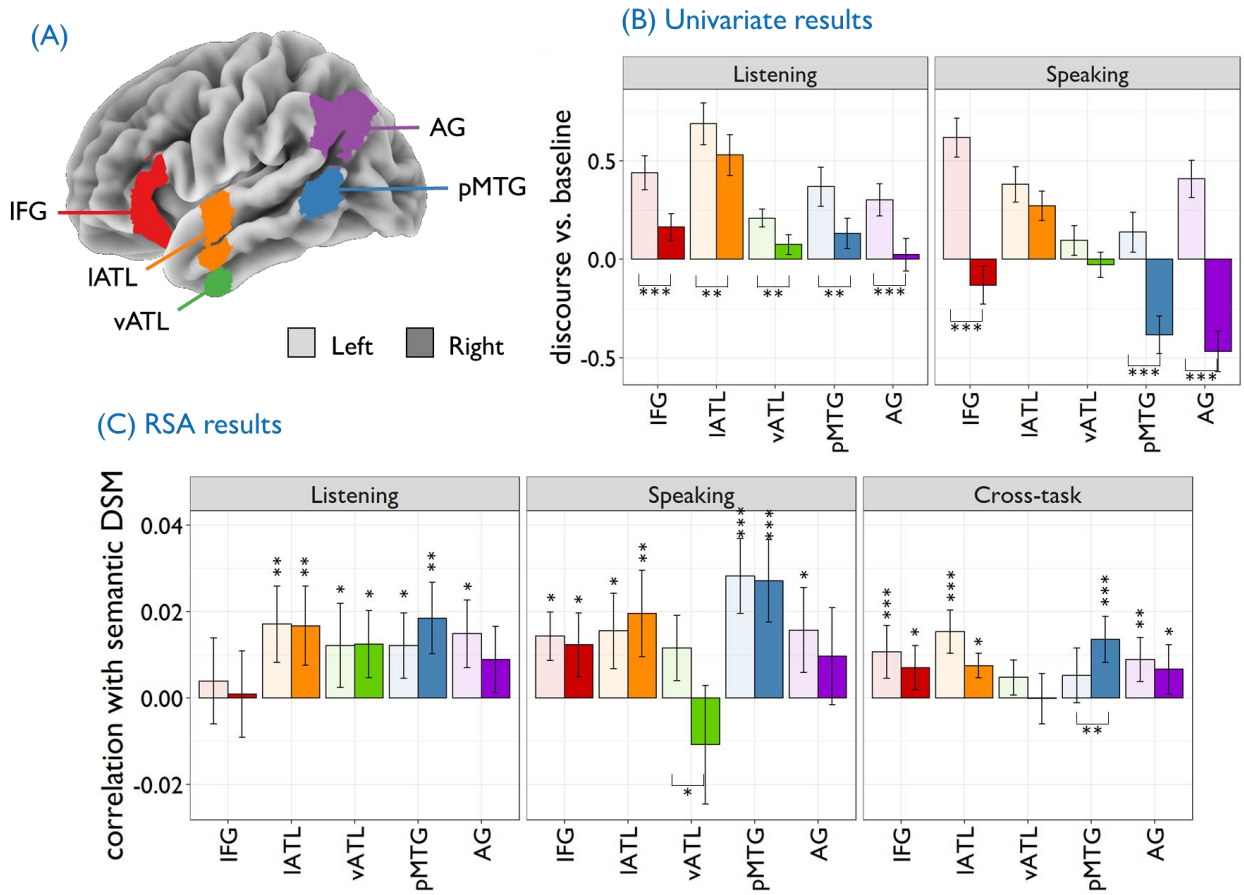


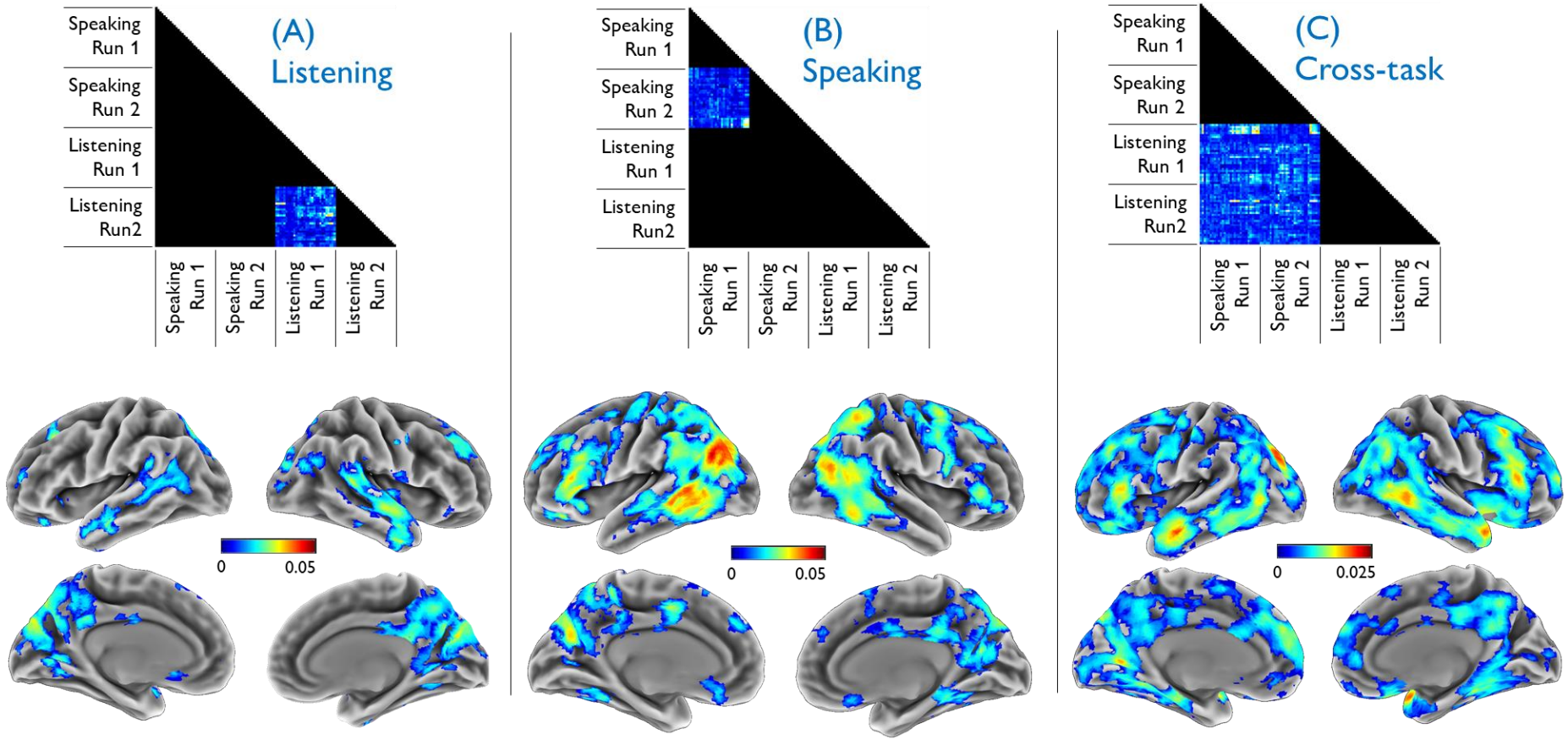
*Figure 3: Region of interest analyses. (A) ROI locations shown on the left hemisphere. Regions were selected based on known involvement in semantic processing and defined using an anatomical atlas (see Method for details). (B) Univariate activation values for discourse vs. baseline, where pale bars represent left-hemisphere activation and bright bars represent the right hemisphere. Asterisks below the x-axis indicate significant hemispheric differences. (C) RSA results, showing the Fisher-transformed correlation between the semantic DSM and neural DSM in each ROI. Asterisks above the x-axis indicate correlations significantly greater than zero, while those below the x-axis indicate significant hemispheric differences. * = p < 0.05; ** = p < 0.01; *** = p < 0.001, all FDR-corrected. Error bars show 1 SEM. IFG = inferior frontal gyrus; lATL = lateral anterior temporal lobe; vATL = ventral anterior temporal lobe; pMTG = posterior middle temporal gyrus; AG = angular gyrus.*

RSA searchlights: These analyses investigated whether similarity in the activation produced during different passages of speech can be explained by similarity in the semantic content of those passages. The RSA method assesses this by first computing a neural dis-similarity matrix (DSM) that measures the dis-similarity (1 – Pearson correlation) between activation patterns elicited at different points during task performance. To generate neural DSMs, we divided each 50s discourse period into 5 segments of 10s duration. We calculated the pairwise dis-

similarities between the activation patterns evoked during these segments of speech (see Methods for details). We then tried to predict the values in the neural DSM using a semantic DSM, which coded dis-similarity between the content of segment using an established vector-based model of semantics [21]. We also controlled for similarity in the quantity of speech contained in each segment (see Figure 1C and Methods). This process was repeated across a series of "searchlights" to build a whole-brain map of semantic-neural correlations. As shown in Figure 4, we performed three separate analyses: comparing listening segments with other listening segments, comparing speaking segments with other speaking segments and, in a cross-task analysis, comparing speaking segments with listening segments.

Results of the three searchlight analyses are shown in Figure 4. When participants listened to discourse, semantic similarity predicted neural similarity in regions along the length of the superior temporal sulci bilaterally. In the anterior temporal lobes, effects extended ventrally into the middle and inferior temporal gyri, particularly in the right hemisphere. Significant correlations were also observed in a large area of posterior medial cortex encompassing posterior cingulate, cuneus and precuneus. These results converge with those of previous studies in indicating that the bilateral lateral temporal lobe regions, as well as other default mode network regions (posterior cingulate, medial prefrontal cortex), track semantic content during spoken narrative comprehension [26, 27].

The second analysis extended the use of semantic RSA to the domain of discourse production for the first time. Here, activity in a more extensive set of regions was found to correlate with semantic similarity (Figure 4B). The strongest correlations were found in left posterior pMTG and inferior parietal cortex, with significant associations also found in IFG, the default mode regions of posterior cingulate and ventromedial prefrontal cortex, as well as parts of the intraparietal sulci and motor cortices. Thus, when individuals produce discourse on similar topics, this is reflected in similar activation patterns within a widely distributed brain network. As in the listening analysis, the distribution of significant regions was largely bilateral, in contrast to the highly left-lateralised pattern observed in the univariate analysis.



<u>Figure 4: Searchlight analyses.</u> The top panel in the figure indicates which parts of the DSMs were used in each analysis. DSMs consist of pairwise comparisons between different 10-second segments of discourse processing. In the Listening and Speaking analyses, we used pairs taken from the same language task but taken from different scanning runs (since segments in the same run may be affected by temporal autocorrelation in the BOLD signal; see Methods). In the cross-task analysis, pairs consisted of one Speaking and one Listening segment (which were always from different scanning runs). Brain maps show regions where the correlation between neural and semantic DSMs was significantly greater than zero (at cluster-corrected $p < 0.05$). Colour scales show the group-average Fisher-transformed correlation coefficient between semantic and neural DSMs; note that the scale is different in the cross-task analysis. Peak effect co-ordinates are reported in Supplementary Table 2.

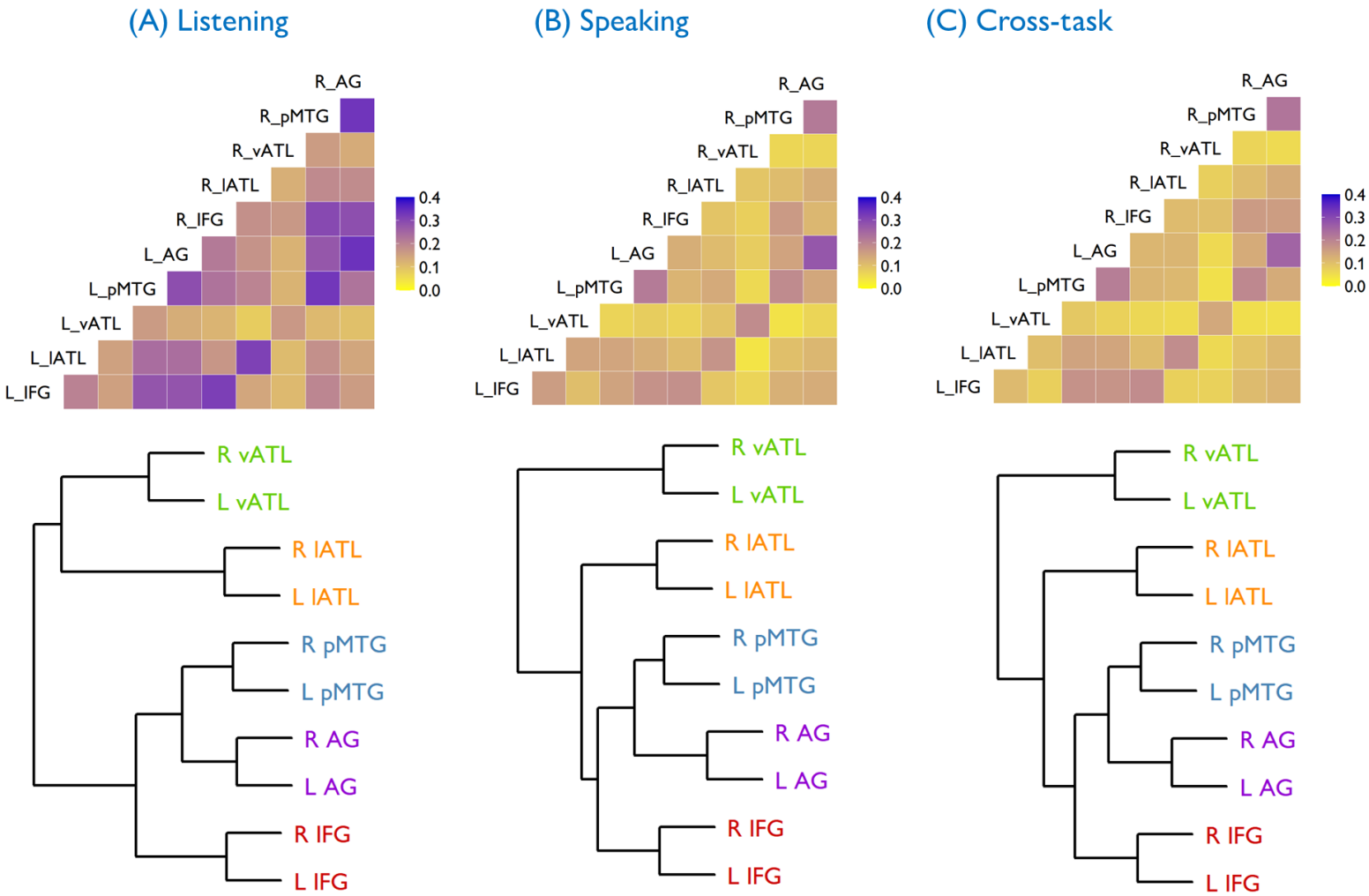


Figure 5: Relationships between neural DSMs. This analysis tested the correlations between neural DSMs obtained in different regions. Top panel shows correlations between neural DSMs for each analysis. Strong correlations indicate that neural DSMs were more similar to one another. Bottom panel shows hierarchical cluster analysis performed using the correlation data. The plots cluster regions according to how similar their neural DSMs are.

Finally, the cross-task analysis tested for predictive effects of semantic similarity when comparing activation patterns across comprehension and production tasks. Significant effects here indicate that similarities in activation across language tasks can be predicted by the underlying semantic content of the discourse being processed. We interpret this as evidence for common coding of language content across speaking and listening tasks. The cross-task analysis revealed semantic-neural correlations in an extensive and bilateral set of regions (Figure 4C). These include regions classically associated with semantic and language processing, such as the anterior and posterior temporal lobes and lateral prefrontal cortices, but also nodes of the default mode network (inferior parietal lobes, posterior cingulate and medial prefrontal cortex) and left hippocampus and parahippocampal gyrus. There are two key implications of these results. First, they indicate that an array of brain regions encode the content of one's own speech in a similar way to the content of speech produced by others, suggesting that the semantics of comprehension and production share a common neural code. Second, they indicate that activation in right-hemisphere regions is as sensitive as left-hemisphere regions to the meaning of discourse. Thus, while univariate analyses suggest that the right hemisphere is less selectively engaged during discourse processing, its activation is nevertheless predicted by the semantic content, suggesting a potential functional role for these regions.

The cross-task analysis identified many of the same regions as the separate task-specific analyses. When comparing these results, however, it is important to note that the magnitudes of the neural-semantic correlations in the cross-task analysis are lower than in the separate listening and speaking analyses. The cross-task analysis had greater statistical power because a greater number of speech segment pairs were available when comparing across different tasks (see Figure 4C). This allowed weaker semantic-neural correlations to reach statistical significance in this analysis.

In the control analyses, neural DSMs were predicted used the quantity DSMs (which class speech segments as dis-similar if they contain very different number of words) rather than semantic DSMs. In the listening task, effects were found bilaterally in ventral premotor and superior temporal regions, centred on Heschl's gyrus (see Supplementary Figure 2). The cross-task analyses revealed similar small clusters in the left and right superior temporal lobes,

again centred on primary auditory cortex. Analysis of the speaking task revealed a more distributed set of regions, including bilateral motor and premotor cortices, some medial parts of the default mode network and left inferior parietal cortex, though with weaker correlations than those seen in the main analyses. None of these analyses revealed the patterns of temporal, inferior prefrontal and parietal effects seen in the main analysis. Thus, the control analyses suggest that effects in Figure 4 are specific to predictors that capture high-level conceptual content of language and not its lower-level properties.

RSA effects in regions of interest: These analyses investigated effects in our set of targeted ROIs that are known to be key nodes in the brain's semantic network [2, 5]. Correlations between semantic and neural DSMs for each ROI are shown in Figure 3C. These are consistent with the widespread and bilateral effects seen in the searchlight analyses. For the listening task, semantic similarity predicted neural similarity in all regions except IFG and right AG (at FDR-corrected $p < 0.05$). For the speaking task, all regions showed significant effects with the exception of vATL and right AG. In the cross-task analysis, significant correlations were observed in all regions except vATL. As with the searchlight results, in almost all cases semantic DSMs predicted neural similarity to a similar degree in the left and right hemispheres. Direct comparisons of the effects in left and right-hemisphere homologues revealed only two significant differences. In the production task, the correlation was significantly stronger in left vATL compared with right vATL. In the cross-task analysis, a significant difference was found in pMTG, but with *right* pMTG showing the stronger effect. Thus, while our univariate ROI analysis showed clearly that semantic regions in the left-hemisphere showed a stronger BOLD response during discourse processing, here we found that activation in both hemispheres tracked the semantic content of discourse to a similar degree.

Results of control ROI analyses, using the quantity DSMs as the predictor, are provided in Supplementary Figure 3. A much more limited set of correlations was observed here. In the listening task, quantity DSMs predicted neural patterns in some regions, predominantly in the temporal lobes (left and right lATL, left vATL and right AG). In the speaking task, only one region (left AG) showed a correlation above zero, and only right IFG showed a cross-task effect. These results suggest that neural patterns in semantic brain

regions were primarily sensitive to the content of discourse and not to the number of words processed in each segment.

Comparison of similarity structure in different brain regions: The final analysis explored the degree to which different parts of the semantic network represent the content of discourse similarly. Rather than comparing each region's neural DSM to a semantic DSM, here we directly compared the neural DSMs of different ROIs with each other (see Figure 1D). The correlations between ROIs for each analysis are shown in the top panel of Figure 5. All of the correlations were positive. This indicates consistency in neural responses across the brain: pairs of speech segments that elicited similar activation patterns in one ROI tended to elicit similar patterns in all of the other ROIs. In general, the strongest correlations between ROIs were observed in the listening task, with less convergence in the speaking task and the cross-task analysis. Strong correlations also tended to be observed between left and right homologue regions (visible as a diagonal line in the lower-right portion of each correlation plot). Correlations between left and right AG were particularly strong. The strong cross-hemispheric coupling in activation is also evident in the hierarchical cluster plots that group ROIs by similarity in their DSMs. In every case, left and right homologues were most similar to one another. This suggests a degree of functional association between left and right-hemisphere regions, despite the fact that the left hemisphere regions consistently showed stronger BOLD responses in the univariate analysis. Otherwise, the three cluster analyses showed broadly similar relationships between regions, with the pMTG and AG most strongly related and also clustered with the IFG, while ATL regions showed more distinct neural patterns.

Discussion

Recent years have seen an explosion in investigations of how language content is represented in the brain, driven by advances in decoding and pattern analysis methods allied with more naturalistic fMRI paradigms [22, 24-27]. Crucially, however, prior studies have

focused almost exclusively on how language is encoded during comprehension tasks, rather than when people produce their own discourse. Here, for the first time, we used a model of semantic discourse content to predict fMRI activation patterns during the *production* of language, comparing this to comprehension-based patterns in the same individuals. We imaged participants while they produced and listened to passages of discourse in response to topic prompts and we used RSA to relate neural similarities between different passages to similarity in their semantic content, using a vector-based model of distributional semantics [21]. When participants produced discourse with similar content, they showed similar patterns of neural activity in a wide range of frontal, temporal and parietal sites, which were more extensive than those observed during comprehension. Moreover, semantic similarity also predicted neural similarities when comparing between the two language tasks, showing that concepts activated during production and comprehension share similar neural representations. Finally, the regions that showed semantic coding of discourse content in both production and comprehension were strikingly bilateral, suggesting a greater role for the right hemisphere in discourse processing than simple activation analyses would suggest.

The major contribution of the present work is to demonstrate similar neural coding of semantic content during production and comprehension of high-level discourse. Previous studies have shown inter-subject correlations when comparing the activation of a speaker and a comprehender [42, 48, 49], and have demonstrated that this coupling is greatest when both participants understand narratives in the same way [54]. The present work makes two important new contributions. First, we show that similar neural patterns for production and comprehension occur within the same individuals as they switch between language tasks. Second, we show that comprehension-production similarities can be predicted by a distributional model of semantics, with semantically similar spoken/heard discourse passages eliciting more similar activation patterns. Alignment between comprehension and production is predicted by a range of theories which propose that higher levels of language processing are largely shared between production and comprehension [35-37]. Potential benefits of this shared cognitive machinery include the ability of the production system to make forward predictions that aid comprehension [38, 55] and the ability for interlocutors to align their mental frameworks during conversation [39].

Semantic effects that crossed between language tasks were observed in a wide range of brain regions. Prominent among these were key nodes of the default-mode network: inferior parietal cortices (AG), posterior cingulate and medial prefrontal regions. Neuroimaging evidence implicates these default-mode network regions in the construction of mental models that represent the details of events (situation models) in a range of contexts, including movie-watching [28], story-listening and reading [42, 54, 56], memory retrieval [57] and in social interactions [53]. Consistent with these findings, our previous work showed increased activation in this network when participants produced or heard discourse that was low in coherence, and thus was harder to construct a situation model for [58]. The present results provide more direct evidence for the role of default-mode regions in representing discourse content, by showing that they represent the semantics of discourse consistently across language processing modes.

We also found effects in ATL regions. Though parts of the ATL are frequently identified as belonging to the default-mode network, theories of semantic cognition typically ascribe a different role to this region compared to other default-mode network regions like AG [59-62]. While AG is thought to be involved in combining concepts to represent events and situations [1, 62, 63], theories of ATL function focus on its role in coding conceptual structure at the single-word and concept level [2, 33]. Our data support this idea that ATL is functionally distinct from AG, since cluster analyses indicated a strong separation between ATL similarity patterns and those of the other semantic regions we studied (including AG but also pMTG and IFG, which we will come to shortly). Nevertheless, there was some evidence that the ATLs were sensitive to the semantic content of speech. Lateral ATLs encoded semantic similarity during both conditions and in the cross-task analysis. The ventral part of the ATL was sensitive to semantic similarity only during speech comprehension. The lack of vATL effects during speech production may be due to this region's high susceptibility to fMRI signal dropout [64].

ROI analyses also revealed effects of semantic similarity in the posterior temporal lobes (pMTG) and prefrontal cortices (IFG). pMTG has been implicated in the semantics of events and actions [65-67], processes which are likely to be critical when understanding or generating narrative discourse. In addition, both pMTG and IFG form part of the semantic

control network: a set of regions which show increased engagement when semantic processing requires high levels of cognitive control [68, 69]. Functions of this network include supporting retrieval of context-relevant semantic information as well as selection processes that arbitrate between multiple competing concepts to ensure contextually relevant information is attended to [2]. Semantic RSA effects in these regions might therefore reflect systematic variations in the cognitive control demands of different discourse topics.

At the whole-brain level, the most striking finding was the widespread right-hemisphere coding of discourse content. This occurred even though right-hemisphere regions were not activated to the same extent as left-hemisphere regions and even, in the case of production, often deactivated relative to baseline conditions or to rest. The cross-ROI analysis further demonstrated that semantic content evoked similar activation patterns across left and right homologues of semantic processing regions. Thus, despite showing different levels of BOLD response, left and right-hemisphere regions appear to track the semantics of discourse in a similar way.

These results are consistent with the view that the right hemisphere makes specific contributions to understanding natural language, such as understanding metaphors and jokes [70, 71], making inferences [72] and comprehending narrative structures [73]. One particular framework suggests that semantic processing occurs bilaterally, with each hemisphere undertaking its own type of neurocomputation but working interactively with the other [74]. The left hemisphere shows fine coding, rapidly activating a network of strongly linked semantic associates; while the right hemisphere encodes more coarse associations, and as such is sensitive to more distant semantic relations and broader conceptual meanings. This coarse coding is thought to be more important for capturing the nuanced meanings contained in more complex language acts, such as extended narratives. Indeed, our findings are consistent with previous studies that have implicated a bilateral network in comprehension of extended, naturalistic speech [e.g., 22]. Importantly, however, we have shown that this right-hemisphere contribution is also present in discourse production. While few neuroimaging studies have investigated language production, lesion studies indicate that patients with right-hemisphere damage produce disorganised and poorly structured narratives, suggesting a right-hemisphere role in discourse planning processes [75-77]. Studies of ATL function also point to a bilateral

pattern of engagement in lexical-semantic processing [78, 79]. Our results are broadly in line with these findings in suggesting that right-hemisphere homologues of semantic regions code the semantic properties of discourse during production as well as comprehension.

We end by considering ways in which the present approach could be extended in future work. First, our semantic model uses information about patterns of word usage in natural language to determine similarity in meaning. Though such distributional approaches can very accurately mimic human judgements of semantic relatedness [80], they have been criticized on the grounds that they do not capture perceptual properties of objects such as their size, shape or colour [81-83]. Future studies may benefit from the use of more complex semantic models that make use of both lexical co-occurrence and experiential attributes [84-87]. Applying such an approach to discourse (cf. single-word comprehension) will present significant challenges, as discourse contains descriptions of complex situations whose perceptual characteristics cannot easily be predicted from their lexical constituents. Second, conceptual representations are dynamic and are shaped by the individual's personal context, in terms of their past experiences, individual processing preferences and abilities [88]. We controlled for these individual differences by directly comparing comprehension and production in the same individuals (as opposed to previous studies which compared speaking and listening in different people). Nevertheless, the degree to which people's discourse representations vary in their content and neural instantiation remains an important question for future work.

To conclude, we have used a distributional model of semantics to predict neural similarity patterns during discourse processing. We have shown that a broad set of brain regions code language content during speech production in a similar way to when hearing someone else's speech, suggesting common coding across different language processing modes. We believe that this work can stimulate further insights into how the semantic systems of the brain drive the generation, as well as understanding, of language.

Method

Participants: 25 young adults participated in the study in exchange for payment. Their mean age was 24 (SD = 4.4, range = 18-35), all were native English speakers and they were all classified as right-handed using the Edinburgh Handedness Inventory [89]. The study was approved by the Psychology Research Ethics Committee of the University of Edinburgh and all participants gave informed consent. Unrelated analyses of this dataset, investigating effects of discourse coherence and other psycholinguistic properties, have been reported previously [58, 90]. In contrast, the present study used multivariate analyses to investigate the neural representation of discourse content.

Materials: In the speaking task, 12 prompts were used to elicit discourse from participants. Each prompt probed common semantic knowledge on a specific topic (e.g., *Describe how you would make a cup of tea or coffee*; see Supplementary Materials for a complete list of prompts in both tasks). This discourse speaking condition was contrasted with a baseline of automatic speech that involved repeated recitation of the English nursery rhyme, Humpty Dumpty [41, 91, 92]. For the listening task, we generated audio samples of discourse using transcripts of speech provided by participants in a previous behavioural study, in response to 12 topic prompts [93]. Listening topics were different from those used in the speaking task. For each topic, we selected two different responses provided by different participants in the Hoffman et al. (2018) study. One sample was highly consistent with the topic and the other deviated somewhat from the specified topic (as indicated by global coherence measured using the methods of Hoffman et al [91, 93]). We did this to increase the variability in the semantic content of the stimuli. We divided the samples into two counterbalanced sets, so that each participant would hear one of the two sets. To generate the audio recordings that would be presented in the scanner, the 24 transcripts were read aloud by the same male native English speaker and edited so that their duration was 50s each (a sample transcript is provided in Supplementary Materials). A 10s recording of the same speaker reciting the Humpty Dumpty rhyme was also made to provide a baseline listening condition.

Design and procedure: Each participant completed two speaking and two listening runs, presented in an alternating sequence. The order of runs was counterbalanced over participants. Each run lasted approximately 8 minutes and included six discourse trials and five baseline trials, with the order of discourse trials in each run fully randomised for each participant.

Speaking trials began with the presentation of a written topic prompt on screen for 6s (see Figure 1A). Participants were asked to prepare to speak during this period and to start speaking when a green circle replaced the prompt in the centre of screen. They were instructed to speak about the topic for 50s, after which a red cross would replace the green circle. At this point participants were instructed to wait for the next prompt to appear on screen. The procedure for listening trials was the same, except participants were asked to listen to the speaker attentively for 50s while the green circle was on screen. For the baseline automatic speech conditions, participants were instructed to recite or to listen to the Humpty Dumpty rhyme for a 10s period. When speaking, they were asked to start reciting again from the beginning if they reached the end of the nursery rhyme before the 10s had elapsed. The baseline conditions therefore involved grammatically well-formed continuous speech, but without the requirement to generate or understand novel discourse. All trials were presented with an inter-stimulus interval jittered between 3s and 7s (M = 5s).

Before scanning, participants were presented with training trials to familiarise them with the tasks. To ensure attention during listening runs, they were told that they would receive a memory test on the material after scanning. In this test, participants answered 12 multiple choice questions, one for each topic presented during listening. They were also asked to rate how well they could hear the speech samples in the scanner from 1 (inaudible) to 7 (perfectly audible).

Processing of speech samples: Spoken responses were digitally recorded with an MRI-compatible microphone and processed with noise cancellation software [94] to reduce noise from the scanner. They were then manually transcribed. For analysis, each 50s response was divided into 5 segments of 10s duration.

For each participant, we constructed two dissimilarity matrices (DSMs) based on the properties of the speech produced or heard in each 10s segment. Our main DSM indexed the semantic relatedness of speech segments using Latent Semantic Analysis (LSA) [21]. Like other distributional models of semantic representation [19, 20], LSA represents words as embeddings in a high-dimensional semantic space. The proximity of two words in this space indicates the degree to which they are used in similar contexts in natural language, which is taken as a proxy for similarity in meaning. We used LSA vectors generated from the British National Corpus, which we have used previously for analyses of discourse production [91, 93]. As in previous studies, a vector representation for each speech segment was generated by averaging the LSA vectors of all of the words contained in the segment [excluding function words and weighting vectors by their log frequency in the segment and their entropy in the original corpus; 93]. Once a vector had been calculated for each speech segment, a semantic DSM was calculated using a cosine similarity metric. An example of a semantic DSM for a single participant is shown in Figure 1B. Each participant's semantic DSM was unique as each participant produced different information in response to the topic prompts.

We computed a second DSM that captured variation in the quantity of speech produced or heard in different segments. Here, dissimilarity between two segments was defined as the difference in the number of words contained in the two segments; thus, segments containing a similar quantity of speech were represented as similar to one another. A quantity DSM for a single participant is shown in Figure 1B. We included the quantity DSM as a covariate in our semantic analyses, to ensure that our effects were being driven by the semantic content of the discourse and were not confounded by variation in the amount of speech participants were processing.

Image acquisition and processing: Participants were scanned on a 3T Siemens Prisma scanner using a 32-channel head coil. fMRI data were acquired at three echo times (13ms, 31ms, and 48ms) using a whole-brain multi-echo acquisition protocol. Data from these three echo series were weighted and combined, and the resulting time-series were denoised using independent components analysis (ICA). This approach improves the signal quality in regions that typically suffer from susceptibility artefacts (e.g., the ventral anterior temporal lobes) and is helpful in

reducing motion-related artefacts [95]. The TR was 1.7 s and images consisted of 46 slices with an 80 x 80 matrix and isotropic voxel size of 3mm. Multiband acceleration with a factor of 2 was used and the flip angle was 73°. Four runs of 281 volumes (477.7s) were acquired. A high-resolution T1-weighted structural image was also acquired for each participant using an MP-RAGE sequence with 1mm isotropic voxels, TR = 2.5 s, TE = 4.6 ms.

Images were pre-processed and analysed using SPM12 and the TE-Dependent Analysis Toolbox 0.0.7 (Tedana) [96]. Estimates of head motion were obtained using the first BOLD echo series. Slice-timing correction was carried out and images were then realigned using the previously obtained motion estimates. Tedana was used to combine the three echo series into a single-time series and to divide the data into components classified as either BOLD-signal or noise-related based on their patterns of signal decay over increasing TEs [95]. Components classified as noise were discarded. After that, images were unwarped with a B0 fieldmap to correct for irregularities in the scanner's magnetic field. Finally, functional images were spatially normalised to MNI space using SPM's DARTEL tool [97].

For univariate analysis, images were smoothed with a kernel of 8mm FWHM. Data were treated with a high-pass filter with a cut-off of 128s and the four experimental runs were analysed using a single general linear model. Four types of speech block were modelled: discourse speaking (50s), baseline speaking (10s), discourse listening (50s), and baseline listening (10s). Prompt presentation periods were modelled as blocks of 6s, with separate regressors for discourse prompts and Humpty Dumpty prompts. Covariates consisted of six motion parameters and their first-order derivatives.

For multivariate analysis, images were smoothed with a kernel of 4mm FWHM, as a small amount of smoothing has been shown to improve the sensitivity of multivariate analyses [98, 99]. Each run was modelled in a separate general linear model. Regressors for prompts and baseline conditions were the same as in the univariate analysis. Discourse blocks were divided into segments of 10s duration (5 per prompt) and each segment was modelled using a separate regressor. Thus, we obtained a separate beta map for each 10s segment of speech heard/produced by each participant. Beta maps were converted into *t*-maps for RSA analysis. Motion covariates were included as in the univariate analysis.

Regions of interest: In addition to whole-brain analyses, we defined five anatomical regions in the left and right hemispheres based on their involvement in semantic cognition. These are shown in Figure 3A. Four of the five ROIs were defined using probability distribution maps from the Harvard-Oxford brain atlas [100], including all voxels with a >30% probability of falling within the following regions:

1. Inferior frontal gyrus (IFG): the pars orbitalis and pars triangularis regions of inferior frontal gyrus, with voxels more medial than x = ±30 removed to exclude medial orbitofrontal cortex
2. Lateral anterior temporal lobe (lATL): the anterior division of the superior and middle temporal gyri
3. Ventral anterior temporal lobe (vATL): the anterior division of the inferior temporal and fusiform gyri
4. Posterior middle temporal gyrus (pMTG): the temporo-occipital part of the middle temporal gyrus

The final ROI covered the angular gyrus (AG) and included voxels with a >30% probability of falling within this region in the LPBA40 atlas [101]. A different atlas was used in this case because the AG region defined in the Harvard-Oxford atlas is small and does not include parts of the inferior parietal cortex typically implicated in semantic processing.

Univariate analyses: We obtained whole-brain activation maps for speech production (discourse minus baseline) and speech comprehension (discourse minus baseline). We computed a subtraction of these two maps to identify differences between speaking and listening at the discourse level. For these analyses we employed a voxel-height threshold of $p < 0.005$ for one-sample *t*-tests with correction for multiple comparisons at the cluster level, using SPM's random field theory. The direct comparison of speaking and listening was restricted to voxels that showed an effect of discourse > baseline in at least one of the two tasks (at a liberal threshold of $p < 0.05$ uncorrected). In addition to whole-brain analysis, contrast estimates for discourse minus baseline were extracted for each ROI in each task and were entered into a 2 x 2 x 5 (task x hemisphere x ROI) ANOVA.

Representational similarity analysis: RSA was used to investigate the degree to which neural patterns were predicted by the semantics of the discourse that participants heard and produced. Analyses were performed using CoSMoMVPA [102]. Searchlight analyses were run using a spherical searchlight of radius 12mm (4 voxels) and ROI analyses were performed in the 10 semantic regions described previously. Because the ROIs varied substantially in size, a voxel selection criterion was applied to ensure that all ROI analyses used the same number of voxels. For each participant, voxels in each ROI were ordered by their effect size in the contrast of discourse over baseline and the 100 most active voxels in each ROI were selected for the RSAs. This ensured that all ROI analyses were equally powered and were based on the voxels that were most engaged by the task.

Separate RSAs were performed for the listening and speaking tasks, as well as a cross-task analysis that tested pattern similarity across comprehension and production. For listening, a neural DSM was generated by calculating 1 minus Pearson correlations between activation patterns for pairs of speech segments presented during the listening task. To avoid any confounding effects of temporal auto-correlation in the BOLD signal, only pairs of segments from different scanning runs were compared [103] (see Figure 4A). This also meant that segments from within the same speech topic were never compared. Activation patterns were mean-centred within each run, ensuring that each voxel had a mean activation of zero [as recommended by 104]. For the speaking task, the same process was followed but pairs were taken from the speaking runs (see Figure 4B). For the cross-task analysis, each pair consisted of one speaking segment and one listening segment (see Figure 4C). Thus, the cross-task analysis tested the degree to which similarity in neural patterns across language tasks could be explained by their semantic content.

The association between neural DSMs and their corresponding semantic DSMs was measured using partial Spearman correlations, which were Fisher-transformed prior to statistical inference. Analyses included the quantity DSMs as a covariate, ensuring that effects were not dependent on the number of words in each segment (see Figure 1C). To test the degree to which our results were specific to semantic processing, control analyses were conducted using quantity DSMs as the predictor and semantic DSMs as the covariate. The results of these are reported in Supplementary Materials.

To determine whether semantic DSMs significantly predicted neural dissimilarity patterns, permutation tests were performed using the two-stage method introduced by Stelzer et al. [105]. For each participant, we calculated Spearman partial correlations using a semantic DSM in which the order of segments had been randomly permuted within each run. This process was repeated 100 times to provide a distribution of results for each participant under the null hypothesis. Following this, a Monte Carlo approach was taken to generate a null distribution at the group level. We randomly selected one correlation map from each participant's null distribution and averaged these to give a group mean. This process was repeated 10,000 times to generate a distribution of the expected group correlations under the null hypothesis. In ROI analyses, the position of the observed result in this null distribution was used to determine a p-value (e.g., if the observed accuracy was greater than 99% of values in the null distribution, this would represent a p-value of 0.01). For searchlight analyses, observed and null maps were entered into CoSMoMVPA's Monte Carlo cluster statistics function, which returned a statistical map corrected for multiple comparisons using threshold-free cluster enhancement [106]. These maps were thresholded at corrected $p < 0.05$.

Comparison of similarity structure in different brain regions: Finally, we investigated the relationship between neural DSMs in different ROIs (see Figure 1D). The purpose of this analysis was to explore the degree to which different parts of the semantic network represented discourse content in a similar way. For each participant, Spearman correlations were calculated between pairs of neural DSMs, for all possible pairs of ROIs. This resulted in a correlation matrix that coded the degree to which pairs of ROIs shared similar neural DSMs. These were then averaged over participants to give a single correlation matrix for the whole group. This process was performed within the listening and speaking tasks, as well as across tasks. To visualise the relationships between ROIs, group correlation matrices were converted to distance matrices and hierarchical cluster analyses performed using R, with Ward's distance metric.

Data and Code Availability: Data and code associated with this study are available at: https://osf.io/s3z8y/. Unthresholded activation and correlation maps are archived at: https://neurovault.org/collections/12792/.

References

1. Binder, J.R. and R.H. Desai, *The neurobiology of semantic memory.* Trends in Cognitive Sciences, 2011. 15(11): p. 527-536.
2. Lambon Ralph, M.A., E. Jefferies, K. Patterson, and T.T. Rogers, *The neural and computational bases of semantic cognition.* Nature Reviews Neuroscience, 2017. 18: p. 42-55.
3. Pulvermüller, F., *How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics.* Trends in cognitive sciences, 2013. 17(9): p. 458-470.
4. Martin, A., *The representation of object concepts in the brain.* Annu Rev Psychol, 2007. 58: p. 25-45.
5. Binder, J.R., R.H. Desai, W.W. Graves, and L.L. Conant, *Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies.* Cerebral Cortex, 2009. 19: p. 2767-2796.
6. Norman, K.A., S.M. Polyn, G.J. Detre, and J.V. Haxby, *Beyond mind-reading: multi-voxel pattern analysis of fMRI data.* Trends in cognitive sciences, 2006. 10(9): p. 424-430.
7. Kriegeskorte, N., M. Mur, and P.A. Bandettini, *Representational similarity analysis-connecting the branches of systems neuroscience.* Frontiers in systems neuroscience, 2008: p. 4.
8. Haxby, J.V., A.C. Connolly, and J.S. Guntupalli, *Decoding neural representational spaces using multivariate pattern analysis.* Annual review of neuroscience, 2014. 37(1): p. 435-456.
9. Wang, X., et al., *Organizational principles of abstract words in the human brain.* Cerebral Cortex, 2018. 28(12): p. 4305-4318.

10. Carota, F., N. Kriegeskorte, H. Nili, and F. Pulvermüller, *Representational similarity mapping of distributional semantics in left inferior frontal, middle temporal, and motor cortex.* Cerebral Cortex, 2017. 27(1): p. 294-309.

11. Devereux, B.J., A. Clarke, A. Marouchos, and L.K. Tyler, *Representational Similarity Analysis Reveals Commonalities and Differences in the Semantic Processing of Words and Objects.* Journal of Neuroscience, 2013. 33(48): p. 18906-18916.

12. Fischer-Baum, S., D. Bruggemann, I.F. Gallego, D.S. Li, and E.R. Tamez, *Decoding levels of representation in reading: A representational similarity approach.* Cortex, 2017. 90: p. 88-102.

13. Mitchell, T.M., S.V. Shinkareva, A. Carlson, K.M. Chang, V.L. Malave, R.A. Mason, and M.A. Just, *Predicting human brain activity associated with the meanings of nouns.* Science, 2008. 320(5880): p. 1191-5.

14. Anderson, A.J., J.R. Binder, L. Fernandino, C.J. Humphries, L.L. Conant, M. Aguilar, X. Wang, D. Doko, and R.D. Raizada, *Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation.* Cerebral Cortex, 2017. 27(9): p. 4379-4395.

15. Pereira, F., B. Lou, B. Pritchett, S. Ritter, S.J. Gershman, N. Kanwisher, M. Botvinick, and E. Fedorenko, *Toward a universal decoder of linguistic meaning from brain activation.* Nature communications, 2018. 9(1): p. 1-13.

16. Wang, J., V.L. Cherkassky, and M.A. Just, *Predicting the brain activation pattern associated with the propositional content of a sentence: modeling neural representations of events and states.* Human brain mapping, 2017. 38(10): p. 4865-4881.

17. Fernandino, L., J.R. Binder, R.H. Desai, S.L. Pendl, C.J. Humphries, W.L. Gross, L.L. Conant, and M.S. Seidenberg, *Concept representation reflects multimodal abstraction: A framework for embodied semantics.* Cerebral cortex, 2016. 26(5): p. 2018-2034.

18. Binder, J.R., L.L. Conant, C.J. Humphries, L. Fernandino, S.B. Simons, M. Aguilar, and R.H. Desai, *Toward a brain-based componential semantic representation.* Cognitive neuropsychology, 2016. 33(3-4): p. 130-174.

19. Mikolov, T., K. Chen, G. Corrado, and J. Dean, *Efficient estimation of word representations in vector space.* arXiv preprint arXiv:1301.3781, 2013.

20. Pennington, J., R. Socher, and C.D. Manning. *Glove: Global vectors for word representation*. in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 2014.

21. Landauer, T.K. and S.T. Dumais, *A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge.* Psychological Review, 1997. 104: p. 211-240.

22. Huth, A.G., W.A. De Heer, T.L. Griffiths, F.E. Theunissen, and J.L. Gallant, *Natural speech reveals the semantic maps that tile human cerebral cortex.* Nature, 2016. 532(7600): p. 453-458.

23. de Heer, W.A., A.G. Huth, T.L. Griffiths, J.L. Gallant, and F.E. Theunissen, *The hierarchical cortical organization of human speech processing.* Journal of Neuroscience, 2017. 37(27): p. 6539-6557.

24. Wehbe, L., B. Murphy, P. Talukdar, A. Fyshe, A. Ramdas, and T. Mitchell, *Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses.* PloS one, 2014. 9(11): p. e112575.

25. Dehghani, M., et al., *Decoding the neural representation of story meanings across languages.* Human brain mapping, 2017. 38(12): p. 6096-6106.

26. Schrimpf, M., I.A. Blank, G. Tuckute, C. Kauf, E.A. Hosseini, N. Kanwisher, J.B. Tenenbaum, and E. Fedorenko, *The neural architecture of language: Integrative modeling converges on predictive processing.* Proceedings of the National Academy of Sciences, 2021. 118(45): p. e2105646118.

27. Zhang, Y., K. Han, R. Worth, and Z. Liu, *Connecting concepts in the brain by mapping cortical representations of semantic relations.* Nature communications, 2020. 11(1): p. 1-13.

28. Baldassano, C., U. Hasson, and K.A. Norman, *Representation of real-world event schemas during narrative perception.* Journal of Neuroscience, 2018. 38(45): p. 9689-9699.

29. Fairhall, S.L. and A. Caramazza, *Brain Regions That Represent Amodal Conceptual Knowledge.* Journal of Neuroscience, 2013. 33(25): p. 10552-10558.

30. Liuzzi, A.G., R. Bruffaerts, R. Peeters, K. Adamczuk, E. Keuleers, S. De Deyne, G. Storms, P. Dupont, and R. Vandenberghe, *Cross-modal representation of spoken and written word meaning in left pars triangularis.* NeuroImage, 2017. 150: p. 292-307.

31. Asyraff, A., R. Lemarchand, A. Tamm, and P. Hoffman, *Stimulus-independent neural coding of event semantics: Evidence from cross-sentence fMRI decoding.* NeuroImage, 2021. 236: p. 118073.

32. Correia, J., E. Formisano, G. Valente, L. Hausfeld, B. Jansma, and M. Bonte, *Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe.* Journal of Neuroscience, 2014. 34(1): p. 332-338.

33. Patterson, K., P.J. Nestor, and T.T. Rogers, *Where do you know what you know? The representation of semantic knowledge in the human brain.* Nature Reviews Neuroscience, 2007. 8(12): p. 976-987.

34. Margulies, D.S., et al., *Situating the default-mode network along a principal gradient of macroscale cortical organization.* Proceedings of the National Academy of Sciences, 2016. 113(44): p. 12574-12579.

35. Kintsch, W. and T.A. Vandijk, *Toward a Model of Text Comprehension and Production.* Psychological Review, 1978. 85(5): p. 363-394.

36. Levelt, W.J., A. Roelofs, and A.S. Meyer, *A theory of lexical access in speech production.* Behavioral and brain sciences, 1999. 22(1): p. 1-38.

37. Hagoort, P., *MUC (Memory, Unification, Control) and beyond.* Front Psychol, 2013. 4: p. 416.

38. Pickering, M.J. and S. Garrod, *An integrated theory of language production and comprehension.* Behav Brain Sci, 2013. 36(4): p. 329-47.

39. Pickering, M.J. and S. Garrod, *Toward a mechanistic psychology of dialogue.* Behavioral and brain sciences, 2004. 27(2): p. 169-190.

40. Awad, M., J.E. Warren, S.K. Scott, F.E. Turkheimer, and R.J. Wise, *A common system for the comprehension and production of narrative speech.* J Neurosci, 2007. 27(43): p. 11455-64.

41. AbdulSabur, N.Y., Y. Xu, S. Liu, H.M. Chow, M. Baxter, J. Carson, and A.R. Braun, *Neural correlates and network connectivity underlying narrative production and comprehension: A combined fMRI and PET study.* Cortex, 2014. 57: p. 107-127.

42. Silbert, L.J., C.J. Honey, E. Simony, D. Poeppel, and U. Hasson, *Coupled neural systems underlie the production and comprehension of naturalistic narrative speech.* Proceedings of the National Academy of Sciences of the United States of America, 2014. 111(43): p. E4687-E4696.

43. Hickok, G. and D. Poeppel, *The cortical organization of speech processing.* Nature Reviews Neuroscience, 2007. 8(5): p. 393-402.

44. Poeppel, D., K. Emmorey, G. Hickok, and L. Pylkkänen, *Towards a new neurobiology of language.* Journal of Neuroscience, 2012. 32(41): p. 14125-14131.

45. Lambon Ralph, M.A., J.L. McClelland, K. Patterson, C.J. Galton, and J.R. Hodges, *No right to speak? The relationship between object naming and semantic impairment: Neuropsychological abstract evidence and a computational model.* Journal of Cognitive Neuroscience, 2001. 13(3): p. 341-356.

46. Schapiro, A.C., J.L. McClelland, S.R. Welbourne, T.T. Rogers, and M.A. Lambon Ralph, *Why bilateral damage is worse than unilateral damage to the brain.* J Cogn Neurosci, 2013. 25(12): p. 2107-23.

47. Federmeier, K.D., *Thinking ahead: The role and roots of prediction in language comprehension.* Psychophysiology, 2007. 44(4): p. 491-505.

48. Stephens, G.J., L.J. Silbert, and U. Hasson, *Speaker-listener neural coupling underlies successful communication.* Proc Natl Acad Sci U S A, 2010. 107(32): p. 14425-30.

49. Liu, Y., E.A. Piazza, E. Simony, P.A. Shewokis, B. Onaral, U. Hasson, and H. Ayaz, *Measuring speaker–listener neural coupling with functional near infrared spectroscopy.* Scientific reports, 2017. 7(1): p. 1-13.

50. Jiang, J., B. Dai, D. Peng, C. Zhu, L. Liu, and C. Lu, *Neural synchronization during face-to-face communication.* Journal of Neuroscience, 2012. 32(45): p. 16064-16069.

51. Badre, D. and A.D. Wagner, *Left ventrolateral prefrontal cortex and the cognitive control of memory.* Neuropsychologia, 2007. 45(13): p. 2883-2901.

52. Jefferies, E., *The neural basis of semantic cognition: Converging evidence from neuropsychology, neuroimaging and TMS.* Cortex, 2013. 49: p. 611-625.

53. Yeshurun, Y., M. Nguyen, and U. Hasson, *The default mode network: where the idiosyncratic self meets the shared social world.* Nature Reviews Neuroscience, 2021. 22(3): p. 181-192.

54. Heidlmayr, K., K. Weber, A. Takashima, and P. Hagoort, *No title, no theme: The joined neural space between speakers and listeners during production and comprehension of multi-sentence discourse.* cortex, 2020. 130: p. 111-126.

55. Dell, G.S. and F. Chang, *The P-chain: Relating sentence production and its disorders to comprehension and acquisition.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2014. 369(1634): p. 20120394.

56. Ferstl, E.C., J. Neumann, C. Bogler, and D.Y. von Cramon, *The extended language network: a meta-analysis of neuroimaging studies on text comprehension.* Hum Brain Mapp, 2008. 29(5): p. 581-93.

57. Ranganath, C. and M. Ritchey, *Two cortical systems for memory-guided behaviour.* Nature reviews neuroscience, 2012. 13(10): p. 713-726.

58. Morales, M., T. Patel, A. Tamm, M. Pickering, and P. Hoffman, *Similar neural networks respond to coherence during comprehension and production of discourse.* Cerebral Cortex, in press.

59. Humphreys, G.F., P. Hoffman, M. Visser, R.J. Binney, and M.A. Lambon Ralph, *Establishing task- and modality-dependent dissociations between the semantic and default mode networks.* Proceedings of the National Academy of Sciences of the United States of America, 2015. 112(25): p. 7857-7862.

60. Farahibozorg, S.-R., R.N. Henson, A.M. Woollams, and O. Hauk, *Distinct roles for the Anterior Temporal Lobe and Angular Gyrus in the spatio-temporal cortical semantic network.* Biorxiv, 2019: p. 544114.

61. Humphreys, G.F., M.A. Lambon Ralph, and J.S. Simons, *A unifying account of angular gyrus contributions to episodic and semantic cognition.* Trends in Neurosciences, 2021. 44(6): p. 452-463.

62. Mirman, D., J.-F. Landrigan, and A.E. Britt, *Taxonomic and thematic semantic systems.* Psychological bulletin, 2017. 143(5): p. 499.

63. Price, A.R., M.F. Bonner, J.E. Peelle, and M. Grossman, *Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus.* The Journal of Neuroscience, 2015. 35(7): p. 3276-3284.

64. Ojemann, J.G., E. Akbudak, A.Z. Snyder, R.C. McKinstry, M.E. Raichle, and T.E. Conturo, *Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts.* Neuroimage, 1997. 6(3): p. 156-167.

65. Davey, J., et al., *Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes.* NeuroImage, 2016. 137: p. 165-177.

66. Liljeström, M., A. Tarkiainen, T. Parviainen, J. Kujala, J. Numminen, J. Hiltunen, M. Laine, and R. Salmelin, *Perceiving and naming actions and objects.* Neuroimage, 2008. 41(3): p. 1132-1141.

67. Watson, C.E., E.R. Cardillo, G.R. Ianni, and A. Chatterjee, *Action concepts in the brain: an activation likelihood estimation meta-analysis.* J Cogn Neurosci, 2013. 25(8): p. 1191-205.

68. Noonan, K.A., E. Jefferies, M. Visser, and M.A. Lambon Ralph, *Going beyond Inferior Prefrontal Involvement in Semantic Control: Evidence for the Additional Contribution of Dorsal Angular Gyrus and Posterior Middle Temporal Cortex.* Journal of Cognitive Neuroscience, 2013. 25(11): p. 1824-1850.

69. Jackson, R.L., *The neural correlates of semantic control revisited.* NeuroImage, 2021. 224: p. 117444.

70. Rapp, A.M., D.E. Mutschler, and M. Erb, *Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies.* Neuroimage, 2012. 63(1): p. 600-610.

71. Marinkovic, K., S. Baldwin, M.G. Courtney, T. Witzel, A.M. Dale, and E. Halgren, *Right hemisphere has the last laugh: neural dynamics of joke appreciation.* Cognitive, Affective, & Behavioral Neuroscience, 2011. 11(1): p. 113-130.

72. Mason, R.A. and M.A. Just, *How the brain processes causal inferences in text: A theoretical account of generation and integration component processes utilizing both cerebral hemispheres.* Psychological Science, 2004. 15(1): p. 1-7.

73. Knutson, K.M., J.N. Wood, and J. Grafman, *Brain activation in processing temporal sequence: an fMRI study.* Neuroimage, 2004. 23(4): p. 1299-1307.

74. Jung-Beeman, M., *Bilateral brain processes for comprehending natural language.* Trends in cognitive sciences, 2005. 9(11): p. 512-518.

75. Marini, A., S. Carlomagno, C. Caltagirone, and U. Nocentini, *The role played by the right hemisphere in the organization of complex textual structures.* Brain and language, 2005. 93(1): p. 46-54.

76. Davis, G.A., T.M. O'Neil-Pirozzi, and M. Coon, *Referential cohesion and logical coherence of narration after right hemisphere stroke.* Brain and Language, 1997. 56(2): p. 183-210.

77. Bartels-Tobin, L.R. and J.J. Hinckley, *Cognition and discourse production in right hemisphere disorder.* Journal of Neurolinguistics, 2005. 18(6): p. 461-477.

78. Rice, G.E., P. Hoffman, and M.A. Lambon Ralph, *Graded specialization within and between the anterior temporal lobes.* Annals of the New York Academy of Sciences, 2015. 1359(1): p. 84-97.

79. Rice, G.E., M.A. Lambon Ralph, and P. Hoffman, *The roles of left versus right anterior temporal lobes in conceptual knowledge: an ALE meta-analysis of 97 functional neuroimaging studies.* Cerebral Cortex, 2015. 25(11): p. 4374-4391.

80. Pereira, F., S. Gershman, S. Ritter, and M. Botvinick, *A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data.* Cognitive neuropsychology, 2016. 33(3-4): p. 175-190.

81. Bruffaerts, R., S. De Deyne, K. Meersmans, A.G. Liuzzi, G. Storms, and R. Vandenberghe, *Redefining the resolution of semantic knowledge in the brain: advances made by the introduction of models of semantics in neuroimaging.* Neuroscience & Biobehavioral Reviews, 2019. 103: p. 3-13.

82. Glenberg, A.M. and D.A. Robertson, *Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning.* Journal of Memory and Language, 2000. 43(3): p. 379-401.

83. Glenberg, A.M. and S. Mehta, *Constraint on covariation: it's not meaning.* Italian Journal of Linguistics. Alessandro Lenci (guest editor): From Context to Meaning: Distributional Models of the Lexicon in Linguistics and Cognitive Science, 2008. 20(1): p. 241-264.

84. Andrews, M., G. Vigliocco, and D. Vinson, *Integrating Experiential and Distributional Data to Learn Semantic Representations.* Psychological Review, 2009. 116(3): p. 463-498.

85. Hoffman, P., J.L. McClelland, and M.A. Lambon Ralph, *Concepts, control and context: A connectionist account of normal and disordered semantic cognition.* Psychological Review, 2018. 125: p. 293-328.

86. Davis, C.P. and E. Yee, *Building semantic memory from embodied and distributional language experience.* Wiley Interdisciplinary Reviews: Cognitive Science, 2021. 12(5): p. e1555.

87. Fernandino, L., J.-Q. Tong, L.L. Conant, C.J. Humphries, and J.R. Binder, *Decoding the information structure underlying the neural representation of concepts.* Proceedings of the National Academy of Sciences, 2022. 119(6): p. e2108091119.

88. Yee, E. and S.L. Thompson-Schill, *Putting concepts into context.* Psychonomic bulletin & review, 2016. 23(4): p. 1015-1027.

89. Oldfield, R.C., *The assessment and analysis of handedness: the Edinburgh inventory.* Neuropsychologia, 1971. 9(1): p. 97-113.

90. Wu, W., M. Morales, T. Patel, M.J. Pickering, and P. Hoffman, *Modulation of brain activity by psycholinguistic information during naturalistic speech comprehension and production.* Cortex, 2022. 155: p. 287-306.

91. Hoffman, P., *Reductions in prefrontal activation predict off-topic utterances during speech production.* Nature communications, 2019. 10(1): p. 515.

92. Blank, S.C., S.K. Scott, K. Murphy, E. Warburton, and R.J. Wise, *Speech production: Wernicke, Broca and beyond.* Brain, 2002. 125(8): p. 1829-1838.

93. Hoffman, P., E. Loginova, and A. Russell, *Poor coherence in older people's speech is explained by impaired semantic and executive processes.* eLife, 2018. 7: p. e38907.

94. Cusack, R., N. Cumming, D. Bor, D. Norris, and J. Lyzenga, *Automated post-hoc noise cancellation tool for audio recordings acquired in an MRI scanner.* Human brain mapping, 2005. 24(4): p. 299-304.

95. Kundu, P., V. Voon, P. Balchandani, M.V. Lombardo, B.A. Poser, and P.A. Bandettini, *Multi-echo fMRI: A review of applications in fMRI denoising and analysis of BOLD signals.* Neuroimage, 2017. 154: p. 59-80.

96. DuPre, E., T. Salo, R. Markello, P. Kundu, K. Whitaker, and D. Handwerker, *ME-ICA/tedana: 0.0.7.* https://doi.org/10.5281/zenodo.3786890, 2019.

97. Ashburner, J., *A fast diffeomorphic image registration algorithm.* Neuroimage, 2007. 38(1): p. 95-113.

98. Hendriks, M.H., N. Daniels, F. Pegado, and H.P. Op de Beeck, *The effect of spatial smoothing on representational similarity in a simple motor paradigm.* Frontiers in neurology, 2017. 8: p. 222.

99. Gardumi, A., D. Ivanov, L. Hausfeld, G. Valente, E. Formisano, and K. Uludağ, *The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis.* Neuroimage, 2016. 132: p. 32-42.

100. Makris, N., J.M. Goldstein, D. Kennedy, S.M. Hodge, V.S. Caviness, S.V. Faraone, M.T. Tsuang, and L.J. Seidman, *Decreased volume of left and total anterior insular lobule in schizophrenia.* Schizophrenia research, 2006. 83(2): p. 155-171.

101. Shattuck, D.W., M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K.L. Narr, R.A. Poldrack, R.M. Bilder, and A.W. Toga, *Construction of a 3D probabilistic atlas of human cortical structures.* Neuroimage, 2008. 39(3): p. 1064-1080.

102. Oosterhof, N.N., A.C. Connolly, and J.V. Haxby, *CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave.* Frontiers in neuroinformatics, 2016. 10: p. 27.

103. Mumford, J.A., T. Davis, and R.A. Poldrack, *The impact of study design on pattern estimation for single-trial multivariate pattern analysis.* Neuroimage, 2014. 103: p. 130-138.

104. Diedrichsen, J. and N. Kriegeskorte, *Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis.* PLoS computational biology, 2017. 13(4): p. e1005508.

105. Stelzer, J., Y. Chen, and R. Turner, *Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control.* Neuroimage, 2013. 65: p. 69-82.

106. Smith, S.M. and T.E. Nichols, *Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference.* Neuroimage, 2009. 44(1): p. 83-98.