The entire brain, more or less is at work: 'Language regions' are artefacts of averaging

Sarah Aliko[1,2], Bangjie Wang[2,3], Steven L Small[4], & Jeremy I Skipper[2*]

[1] London Interdisciplinary Biosciences Consortium, University College London, UK

[2] Experimental Psychology, University College London, UK

[3] Department of Psychology, School of Behavioral and Brain Sciences, University of Texas at Dallas, USA

[4] School of Behavioral and Brain Sciences, University of Texas at Dallas, USA

[*] Corresponding author: jeremy.skipper@ucl.ac.uk

Abstract

Models of the neurobiology of language suggest that a small number of anatomically fixed brain regions are responsible for language functioning. This observation derives from centuries of examining brain lesions causing aphasia and is supported by decades of neuroimaging studies. The latter rely on thresholded measures of central tendency applied to activity patterns resulting from heterogeneous stimuli. We hypothesised that these methods obscure the whole brain distribution of regions supporting language. Specifically, cortical 'language regions' and the corresponding 'language network' consist of input regions and connectivity hubs. The latter primarily coordinate peripheral regions whose activity is variable, making them likely to be averaged out following thresholding. We tested these hypotheses in two studies using neuroimaging meta-analyses and functional magnetic resonance imaging during film watching. Both converged to suggest that averaging over heterogeneous words is localised to regions historically associated with language but distributed throughout most of the brain when not averaging over the sensorimotor properties of those words. The localised word regions are composed of highly central hubs. The film data further shows that these hubs are dynamic, connected to peripheral regions, and only appear in the aggregate across time. Results suggest that 'language regions' are an artefact of indiscriminately averaging across heterogeneous language representations and linguistic processes. Rather, they are mostly dynamic connectivity hubs coordinating whole-brain distributions of networks for processing the complexities of real-world language use, explaining why damage to them results in aphasia.

Keywords. Averaging; Broca's Area; Centrality; Connectivity; fMRI; Hubs; Language Network; Language Regions; Neurobiology of Language; Wernicke's Area

Beware of the Blob! It creeps, and leaps, and glides and slides...

Burt Bacharach, 1958

Introduction

Historical views of ‘the organisation of language and the brain’ derive from mid-19th century post-mortem lesion analyses. For the next 100 years, there existed an ongoing debate as to whether language is localised to specific brain regions or holistic, involving the whole brain. The holistic view lost ground in the 1960s when Norman Geschwind helped swing the pendulum back to a localizationist model (1). Roughly, this model posits that a left posterior superior temporal lobe (or ‘Wernicke’s area’; but see)(2) and left inferior frontal gyrus (or ‘Broca’s area’) are the anatomical loci for language comprehension and speech production, respectively (3). These localisations and ‘localizationism’ have subsequently remained the standard view, likely for a number of reasons. Most prominently, beginning in the 1970s and continuing until now, tens of thousands of in vivo positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies of language appear to support these anatomical observations.

Advances from neuroimaging evidence have arguably been only marginally incremental over Geschwind’s now ‘classical’ model. Specifically, contemporary models of the systems neurobiology of language recognise that ‘language’ is a complex process that can be decomposed into subprocesses that are localised to a few additional regions. These are all in close proximity to those in the classical model or their right hemisphere ‘homologues’ but not typically ascribed to that older model. They include more of the superior temporal lobe, as anterior as the temporal pole, middle temporal lobe, and premotor cortices (4–7). This neo-localizationist view is communicated in the scientific literature using phrases like ‘language regions’ to describe portions of the brain and ‘the language network’ to describe commonly co-active regions of the brain[1]. Indeed, these conceptualizations seem to be validated by the anatomical regions emerging from neuroimaging meta-analyses, conducted across a wide variety of language stimuli and tasks (see Figure 1).

This apparent consistency begs the question as to why those specific anatomical locations support speech perception and language comprehension and why damage to them results in aphasia. One obvious answer is that they mostly involve regions adjacent to those in which acoustic information from the cochlea first arrives in the neocortex. It stands to reason that nearby regions would act upon that auditory information and transform it into the putative representations and processes that

[1] ‘Language regions’, ‘the language network’, and ‘language network’ result in more than 12,000, 9,000, and 23,000 links on Google Scholar, respectively (ascertained September, 2023).

distinguish mere sound from language (like phonemes and phonology). It would be inefficient for distant regions to contribute because of the time delays and metabolic costs. Indeed, studies using ‘language localisers’ suggest that ‘the language network’ processes only language information whereas more broadly distributed regions perform ‘extraneous’ and ‘autonomous’ functions ‘not necessary for linguistic processing’ (8–11).

Averaging

An alternative model of the neurobiology of language explains the apparent anatomical consistency that is reflected by classical and contemporary models differently (12). It starts with the behavioural observation that ‘language’ is not only complex but also ambiguous at all levels, from speech sounds to semantics, syntax, and discourse. To resolve this ‘problem’, a large body of empirical evidence suggests that the brain makes use of contextual information stored internally (e.g., in the form of knowledge and expectations) and available externally (e.g., in the form of observable speech associated mouth movements and co-speech gestures)(13–16)(for further examples, see)(12). The memory and perceptual processes associated with internal and external context are distributed across the whole brain (13–19). These distributed networks appear to predict language input, e.g., in primary auditory cortex (20). Because context varies dynamically with each linguistic experience, language processing might never be the same twice (12, 21). As such, the distributed regions involved in predicting language input would themselves be highly variable and dynamic.

Such a model is in the unenviable position of not conforming to Occam's razor. That is, in addition to having to explain the consistency of regions appearing across studies (see Figure 1), it must explain why we do not typically observe a whole-brain distribution of other regions engaged in language processing. One explanation is that they are concealed by existing methodological paradigms. For example, all neuroimaging studies rely on measures of central tendency, like averaging. In context-aware neurobiological models of language like that outlined in the prior paragraph, every word has distributed and variable activity patterns. Thus, averaging over different words, whether individually or in n-grams, sentences, or discourse, would reduce the activity in those ‘peripheral’ regions with variable activity patterns to a low value.

This low value is even less likely to survive given the specific statistical analyses used. To illustrate, participants might listen to 100 different words during fMRI. The resulting data is a four-dimensional matrix of tens of thousands of ‘voxels’ collected at multiple time steps. Regressions are conducted in each of those voxels at the individual participant level, using a regressor that is a convolution of word onset times with a ‘canonical’ hemodynamic response function. Though this is not technically an average, it effectively acts as one by collapsing over different word representations. Resulting

regression coefficients are then used to conduct group level statistical analyses, again, at each voxel, resulting in average coefficients across participants. This is yet another level of collapsing over the different representations associated with each word as these may differ by participant. Finally, if any voxels exceed a statistical 'threshold' it is said that they are 'activated'. However, thresholding requires prohibitive corrections for multiple comparisons because of the number of statistical tests that are done[2]. Thus, connected peripheral voxels with legitimate but variable activity for any individual word are statistically unlikely to survive collapsing and averaging over word representations at the individual and group levels, particularly after thresholding. This would give the false impression that those vowels are inactive and, therefore, 'extraneous'[3].

Distributed

Consistent with this picture, there is a large amount of evidence that a distributed context-aware model of the neurobiology of language is more consistent with reality when analyses do not indiscriminately average over heterogeneous word representations. To give an example, word processing can be considered to be instantiated in a distributed neural ensemble that incorporates experiences associated with word learning (12, 22). Within this framework, verbs activate brain regions responsible for motion perception and limb movement whereas nouns are more associated with object processing regions (23). This is also true of finer representations, with words having auditory, visual, somatosensory, movement, and emotional related meanings activating regions associated with those processes, e.g., the transverse temporal gyrus, calcarine sulcus, postcentral gyrus, central sulcus, and insula, respectively (24–26). One cannot argue that these patterns simply represent post-linguistic 'conceptual' processing as they occur before this is possible, 50-150 ms after word onset (27–31).

Results like these suggest that language processing is distributed across the whole brain, most of which is obscured by averaging words from different semantic categories. A similar conclusion can be drawn from other linguistic representations. For example, overlearned 'formulaic' language, e.g. 'you know' seems to not even involve classical 'language regions', explaining why they are often preserved after severe damage, e.g., as in global aphasia (32, 33). A similar argument can be made not only for linguistic representations but other linguistic processes. For example, syntactic processing is arguably distributed (34), with different brain regions participating in different syntactic functions when not

---

[2] For example, to Bonferroni correct a single 'mass univariate' statistical image with two mm^3 voxels, the p-value should be 0.05/250,000 or $p \leq 0.0000002$. That said, corrections are often not this stringent because allowances are made for the fact that voxels are not remotely independent.

[3] Thus, the word 'averaging' as used throughout is a convenient simplification for this particular set of analyses steps and other similar neuroimaging analyses. Arguably, all neuroimaging studies involve measures of central tendency or averaging in this sense.

collapsed over (35). These include regions outside of the classical and contemporary ‘language regions’, e.g., the basal ganglia, pre-supplementary motor area, and insula (36, 37).

Compounding the issue, averaging over all of these linguistic representations and processes typically conceals the distributed patterns associated with individual differences. For example, left and right handed participants activate right and left motor cortices, respectively, for verbs like ‘throw’ (38). Similarly, professional hockey players activate motor cortices for hockey related sentences compared to fans and novices (39, 40). More generally, we have known for more than 20 years that individual differences in patterns of language associated activity do not correspond well to group averages (41). Group analysis by clustering shows extensive variability during language comprehension, with no one group of participants capturing the aggregate and individual participants varying on a spectrum involving the relative contribution of multiple neural structures, e.g., visual and sensorimotor regions (42).

Hubs

We return to the question of why ‘language regions’ or ‘the language network’ remain after averaging if standard methodological practices obscure most of the participating brain regions? One suggestion derives from the fact that the network organisation of the brain is neither random or uniform. Rather, it has a small-world topology, characterised by the presence of hubs (43). The latter are regions that are highly central (i.e., have a high degree), with a large number of connections to other regions (44). This leads to the hypothesis that ‘language regions’ are a combination of auditory regions due to persistent audio input (at least in spoken language) and hubs necessary for the coordination of information processing in more dynamic, variable, and distributed peripheral regions.

Indeed, empirical evidence hints at the possibility that the regions identified in classical and contemporary models of the neurobiology of language are connectivity hubs. Structural MRI (45–47) and ‘resting-state’ fMRI studies (48–58) suggest that portions of superior and middle temporal and inferior frontal gyri are hubs. Though there are few studies measuring hubs using auditory or language stimuli, those that exist again suggest that these regions are functional hubs, encompassing even ‘early’ auditory cortices (59, 60). However, across all structural, resting-state, and task-based studies, the location of hubs in this set of language associated regions is quite variable. For example, hubs variously encompass portions of the anterior, middle, or posterior superior temporal gyrus and sulcus, suggesting they might be dynamic (61).

Hypotheses and Studies

To summarise, there is a conflict between the localizationist claims about ‘language regions’ and ‘the language network’ and data demonstrating that language processing occurs in a widely distributed set of brain regions. An alternative to localizationism is that these networks are not more obvious because averaging (and thresholding) over the more variable and distributed regions leaves only auditory input regions and connectivity hubs that coordinate those more distributed peripheral regions. As we are not aware of any work empirically demonstrating this, we conducted two neuroimaging based studies to test the following four hypotheses:

H1-Averaging	*Averages over disparate linguistic representations and processes are localised to ‘language regions’.*
H2-Distributed	*When not averaged over, specific linguistic representations and processes are distributed throughout the whole brain.*
H3-Hubs	*The ‘language regions’ that survive averaging are auditory input regions and connectivity hubs.*
H4-Dynamics	*Language associated connectivity hubs are dynamic, not fixed, and appear only in the aggregate.*

Meta-Analyses

We test the first hypothesis by conducting a meta-analysis of neuroimaging meta-analyses or a ‘meta-meta-analysis’ of language (62). In doing so, we attempt to demonstrate that studies that average over a variety of linguistic representations and processes consistently show activity in the same ‘language regions’ (H1-Averaging). However, this would not adjudicate between the veracity of the localizationist view versus the view that averaging conceals distributed regions. For this, we perform a second set of meta-analyses to determine if language processing in the brain is distributed when comparing even gross linguistic representations like verbs and nouns (H2-Distributed). Finally, we conduct a meta-analytic centrality analysis to test the third hypothesis that the repeatedly activated ‘language regions’ expected in the meta-meta-analysis are hubs (H3-Hubs).

NNDb

The preceding meta-analyses can serve as only an approximate validation of the hypotheses 1-3. That is, even if the meta-meta-analysis reveals only ‘language regions’ and the verbs and nouns meta-analyses reveal a distributed pattern, the former cannot be derived by averaging over the latter. This and determining the relationship between ‘language regions’ and connectivity hubs must also be demonstrated within participants. Furthermore, meta-analyses are static and cannot say anything about the dynamics of activity (required for H4-Dynamics). Thus, we used fMRI in a second study to test

hypotheses within participants using our ‘naturalistic neuroimaging database’ (NNDb)(63). Specifically, we test our hypotheses that language processing appears circumscribed to ‘language regions’ when using a measure of central tendency over words in the films participants watched (H1-Averaging) but widely distributed when examining finer sensorimotor representations of those words (H2-Distributed). We tested the third hypothesis by calculating the extent to which these averaged word regions are hubs, using multiple measures of voxel-wise network centrality (H3-Hubs).

We specifically used the NNDb to be able to test our fourth hypothesis, which derives from the context-aware model discussed above (see ‘Averaging’), requiring a stimulus that is temporally extended and context rich. In particular, our model proposes that the forms of context used during naturalistic language processing vary dynamically and are therefore coordinated by spatially and temporally varying hubs. For example, at one moment speech-associated mouth movements might be used to predict forthcoming speech sounds, involving more posterior superior temporal hubs whereas at another moment iconic co-speech gestures might be used to predict upcoming words, involving more anterior superior temporal hubs (13–16, 20). As such, we hypothesised that coordinating hubs are not a fixed set that always appear together when language is being processed. Rather, what have come to be called ‘language regions’ or ‘the language network’ are a virtual set that appears only collectively, when aggregating over an extended period of processing time (H4-Dynamics). To be able to test this hypothesis, we derived our measures of voxel-wise network centrality using dynamic functional connectivity at overlapping running windows so that we could perform analysis both with and without aggregative over entire films.

Results

Meta-Analyses

Averaging

We first conducted a neuroimaging ‘meta-meta-analysis’ to localise ‘language regions’ and to determine if putatively different language representations and processes activate similar ‘language regions’ across different meta-analyses. Such a result would be expected both under the hypothesis that ‘language regions’ are a self-contained localised system or network and the hypothesis that they are primarily a product of averaging. Specifically, 85 language related meta-analyses were conducted (see Table 1, Materials and Methods) and thresholded, correcting for multiple comparisons. Results were combined into one neuroimage by count, followed by a one-sample ‘group’ level statistical analysis, again correcting for multiple comparisons.

The results of this analysis revealed a circumscribed distribution of voxels activated in one or more language related meta-analyses (Figure 1). On average, each voxel was activated in 6.36 of the

meta-analyses, with a maximum single voxel activation in 45 meta-analyses (in the left dorsal superior temporal sulcus). Significant temporal lobe regions were activated across meta-analyses in both hemispheres, including the transverse temporal gyri, plana polare and temporale, superior temporal gyri and sulci, and posterior middle temporal gyri (Figure 1, white outline). Significant activity only in the left hemisphere included the posterior inferior frontal gyrus, ventral and dorsal precentral sulcus and gyrus, and medial superior frontal gyrus (Figure 1, white outline).

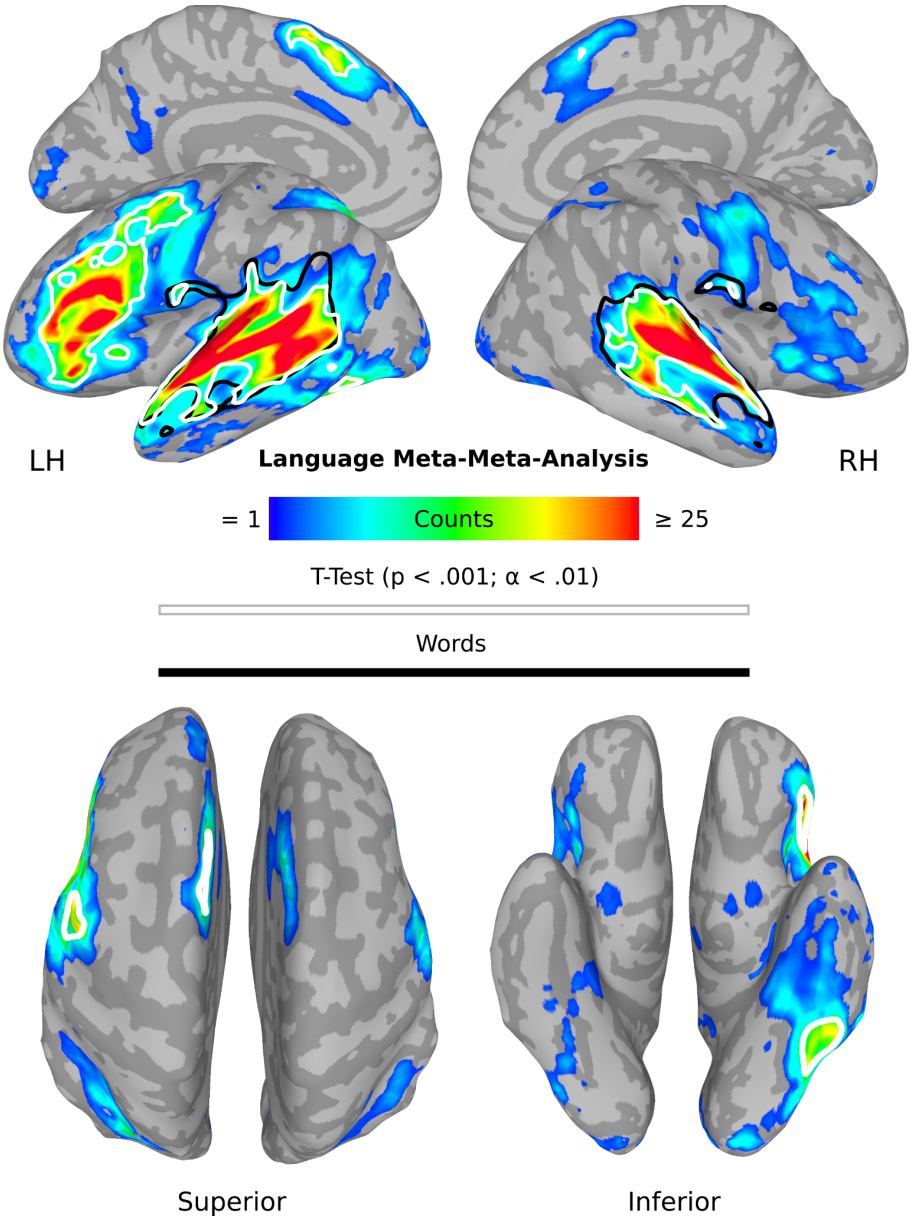

Figure 1. *Neuroimaging meta-meta-analysis of language. Eighty-five meta-analyses of language representations (e.g., phonemes, words, sentences) and associated processes (e.g., speech, semantics, syntax) were conducted using the Neurosynth (N = 57) and BrainMap (N = 28) databases. Each meta-analysis was thresholded at p ≤ 0.01 false discovery rate corrected for multiple comparisons using a minimum cluster size of 50 voxels (400 ml). This left 45 meta-analyses that each activated*

approximately the same regions. This can be gleaned from the number of meta-analyses active in each voxel (colours) and a t-test (using database as a covariate and cluster size corrected; white outline). The main effect of words from the NNDb study is shown with a black outline for comparison (See Figures 4 and 5).

Distributed

Second, we performed neuroimaging meta-analyses of ‘verbs’ and ‘nouns’ as an example to demonstrate that language is distributed across the whole brain when more specific linguistic representations are not averaged over, as is typically the case. This would serve as evidence that the ‘language regions’ in the meta-meta-analysis are in fact more likely to be the result of averaging. Indeed, verbs and nouns produce a whole brain distribution of activity (Figure 2, colours) that is not encompassed by the significant ‘language regions’ from the meta-meta-analysis (Figure 2, white outline). Specifically, significant activity for verbs included bilateral activity in the dorsal precentral sulcus and gyrus, central sulcus, postcentral gyrus, medial superior frontal gyrus, and medial temporal area (among other regions; Figure 2, red). In contrast, nouns produced activity in bilateral occipital cortex, fusiform gyrus, and inferior occipital gyrus continuing into the inferior temporal gyrus (among other regions; Figure 2, blue).

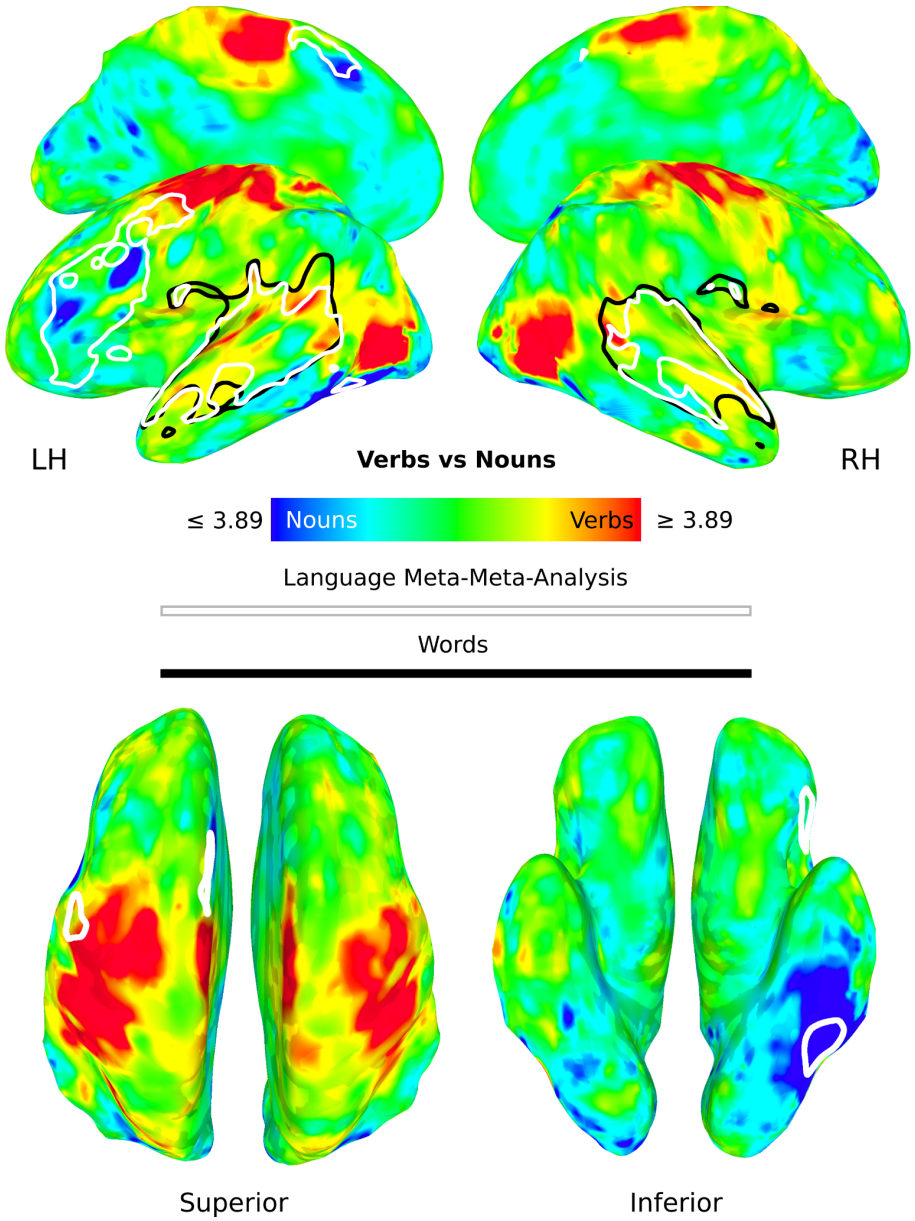

Figure 2. *Neuroimaging meta-analysis of verbs and nouns. A meta-analysis of verbs (N = 662) and nouns (N = 889) was conducted using the NeuroQuery database. The z-scores were subtracted and presented unthresholded. Dark red and blue regions represent verbs and nouns, respectively, at a difference in the absolute value of the z-scores ≥ 3.89 or p ≤ 0.0001. Significant voxels from the language meta-meta-analysis in Figure 1 are shown as a white outline. The main effect of words from the NNDb fMRI study is shown with a black outline for comparison (See Figures 4 and 5).*

Hubs

Finally, we developed a new meta-analytic approach to quantifying degree centrality in order to evaluate whether or not the ‘language regions’ in the meta-meta-analysis are network connectivity hubs (64, 65). If they are hubs, this would suggest that ‘language regions’ in the meta-meta-analysis

remain after averaging over heterogeneous linguistic representations (like verbs and nouns) because they are hubs coordinating more distributed linguistic representations and processes. Specifically, we conduct 165,953 voxel-wise co-activation meta-analyses across 14,371 studies. As with functional connectivity, the regular co-activation of two or more regions suggests that those regions form functional connections or a network. Each meta-analysis was thresholded and combined by count (i.e., degree) as a measure of centrality and converted to a z-score for later thresholding.

The results indicate that a circumscribed set of regions in the task-driven brain have high centrality (Figure 3). These include many regions in the superior and middle temporal lobes, posterior inferior frontal, and precentral regions from the meta-meta-analysis (Figure 3, white outline). Indeed, the spatial correlation between the unthresholded centrality and language meta-meta-analysis was $r = 0.52$. When the centrality results were thresholded at $z = 3.89$ ($p \leq 0.0001$), $z = 3.29$ ($p \leq 0.001$), $z = 2.58$ ($p \leq 0.01$), or $z = 1.96$ ($p \leq 0.05$), they occupied 34%, 45%, 62%, and 76% of thresholded language meta-meta-analysis voxels, respectively.

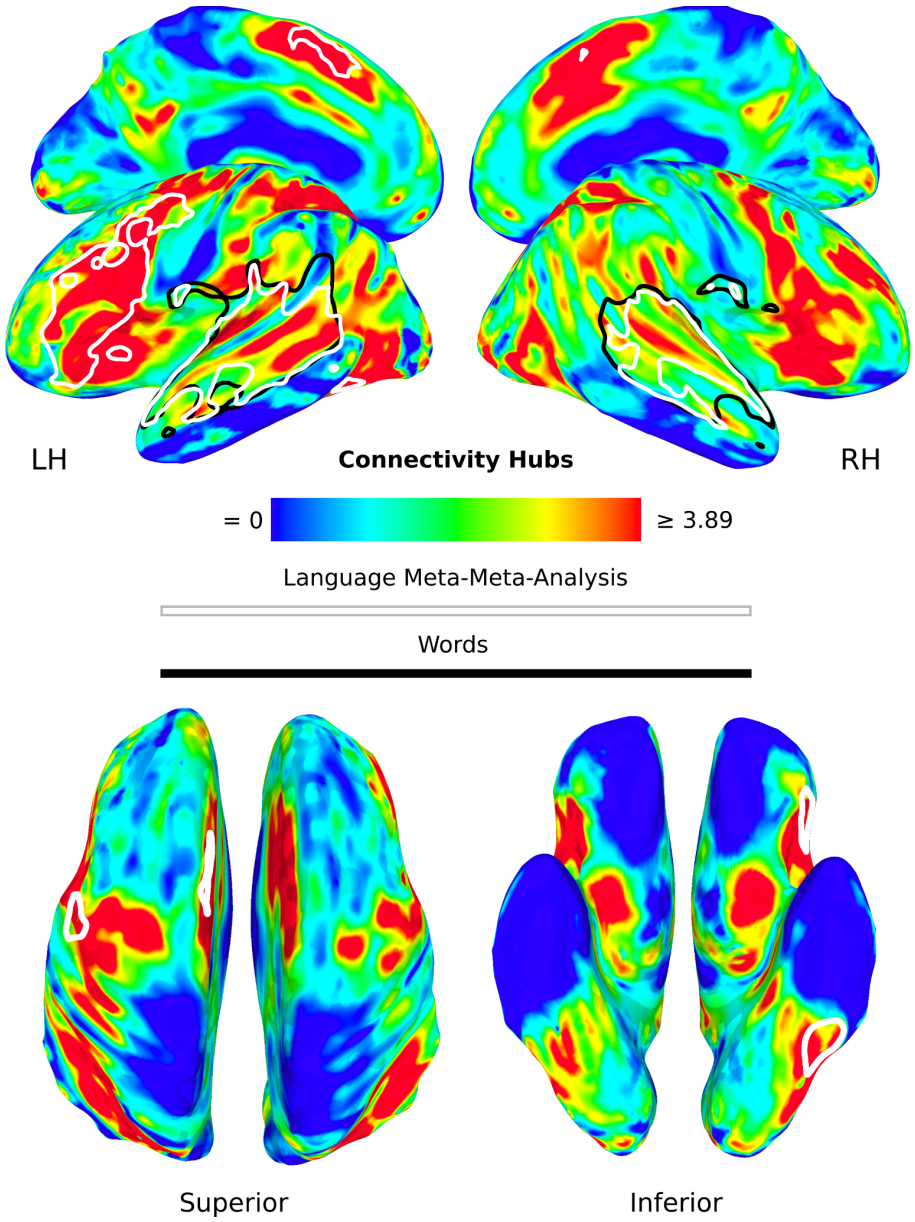

Figure 3. *Neuroimaging meta-analytic connectivity hubs. Degree centrality was calculated using 165,953 voxel-wise co-activation meta-analyses across 14,371 studies, thresholded at $p \leq 0.01$ false discovery rate corrected for multiple comparisons and combined by count. The mean and standard deviation across grey matter were used to convert counts to z-scores to make interpretation easier. Dark red regions are those that have z-scores ≥ 3.89 or $p \leq 0.0001$. Significant voxels from the language meta-meta-analysis in Figure 1 are shown as a white outline. The main effect of words from the NNDb fMRI study is shown with a black outline for comparison (See Figures 4 and 5).*

NNDb

The results from the neuroimaging meta-analyses suggest that ‘language regions’ and ‘the language network’ might be the product of averaging over different word representations (Figures 1-2), leaving

only connectivity hubs (Figure 3). However, the results are only suggestive because neither averaging nor connectivity hubs were generated within any one study, meaning results might be due to some factor unrelated to our hypotheses.

Averaging

To overcome this issue and for other reasons discussed in the Introduction (see Hypotheses and Studies), we used film-fMRI data from 86 participants to average over the heterogeneous properties of words in the brain, with the expectation that the results will be similar to the language meta-meta-analysis (Figure 1). Specifically, we used duration and amplitude modulated regression at the individual participant level, modelling words and allowing them to be modulated by 11 sensorimotor experiential dimensions associated with the meanings of those words (i.e., auditory, foot-leg, gustatory, hand-arm, haptic, head, interoceptive, mouth, olfactory, torso, and visual dimensions). We also included sound energy, contrast luminance, and word frequency as nuisance modulators and regressors for words without sensorimotor ratings and nonword film segments. For the group level analysis, the beta maps corresponding to the ‘main effect’ of words (henceforth ‘words’) and the 11 sensorimotor modulators were entered into a linear mixed-effects model with beta, age, gender, and film as fixed effects and participant as a random effect. The results were corrected for multiple comparisons using a cluster-size correction with multiple voxel-wise thresholds to $\alpha = 0.01$ and a minimal cluster of 20 voxels (540 ml).

Consistent with our hypothesis, word processing in the brain was spatially limited (Figure 4). Regions of activity for the words included the same bilateral superior and middle temporal lobe regions as in the meta-meta-analysis (Figure 4, black and white outlines). The percentage of thresholded word voxels in the language meta-meta-analysis was 53% (1414 out of 2686 voxels). The predominant difference between these maps was the lack of the left inferior frontal and precentral activity (though some of this, i.e., the pars opercularis and precentral gyrus and sulcus are more active for word contrasts, not presented). In contrast, the percentage of 11 thresholded sensorimotor modulators of words (e.g, hand/arm) in the language meta-meta-analysis was 7% (M = 199.27 of 2686 voxels, SD = 380.71; discussed further in the next section). The spatial pattern of activity for the unthresholded words map was correlated with the unthresholded language meta-meta-analysis at $r = 0.51$. In contrast, the 11 sensorimotor modulators of words did not correlate with the language meta-meta-analysis on average (M r = -0.01, SD = 0.13).

Distributed

In parallel with the verb/noun meta-analyses (Figure 2), we analysed the fine scale sensorimotor properties of words within participants, evaluating the regional distribution in the brain. In this

analysis, sensorimotor processing was distributed throughout much of the brain (Figure 4, yellows and reds). Figure 4 shows the main effect of the 11 sensorimotor modulators (e.g., foot/leg, Figure 4, reds) and pairwise contrasts between modulators (e.g, foot/leg vs hand/arm, Figure 4, yellows) averaged and projected onto the brain, along with voxels that were not significant in any of these comparisons (Figure 4, blues). In total, 37% of the brain was activated by sensorimotor modulators outside of voxels activated by words (17,718 of 48,581 voxels), with each modulator activating an average of 4% of those voxels (2,098.27 of 48,581 voxels, SD = 2,599.19). Similarly, when considering direct contrasts between modulators, 67% of the brain was activated outside of word voxels (32,489 of 48,581 voxels), with each contrast again activating an average of 4% of those voxels (2,095.71 of 48,581 voxels, SD = 1,839.69). For comparison, words activated about 5% of the whole brain in total (2,328 of 50,909 voxels). Finally, neither the spatial pattern of activity for the unthresholded sensorimotor (mean r = -0.02, SD = 0.24) nor the direct contrasts maps (mean r = -0.05, SD = 0.22) were correlated with the unthresholded word map on average.[4]

[4] Note that all calculations done in this paragraph are similar when done with the language meta-meta-analysis instead of the main effect of words.

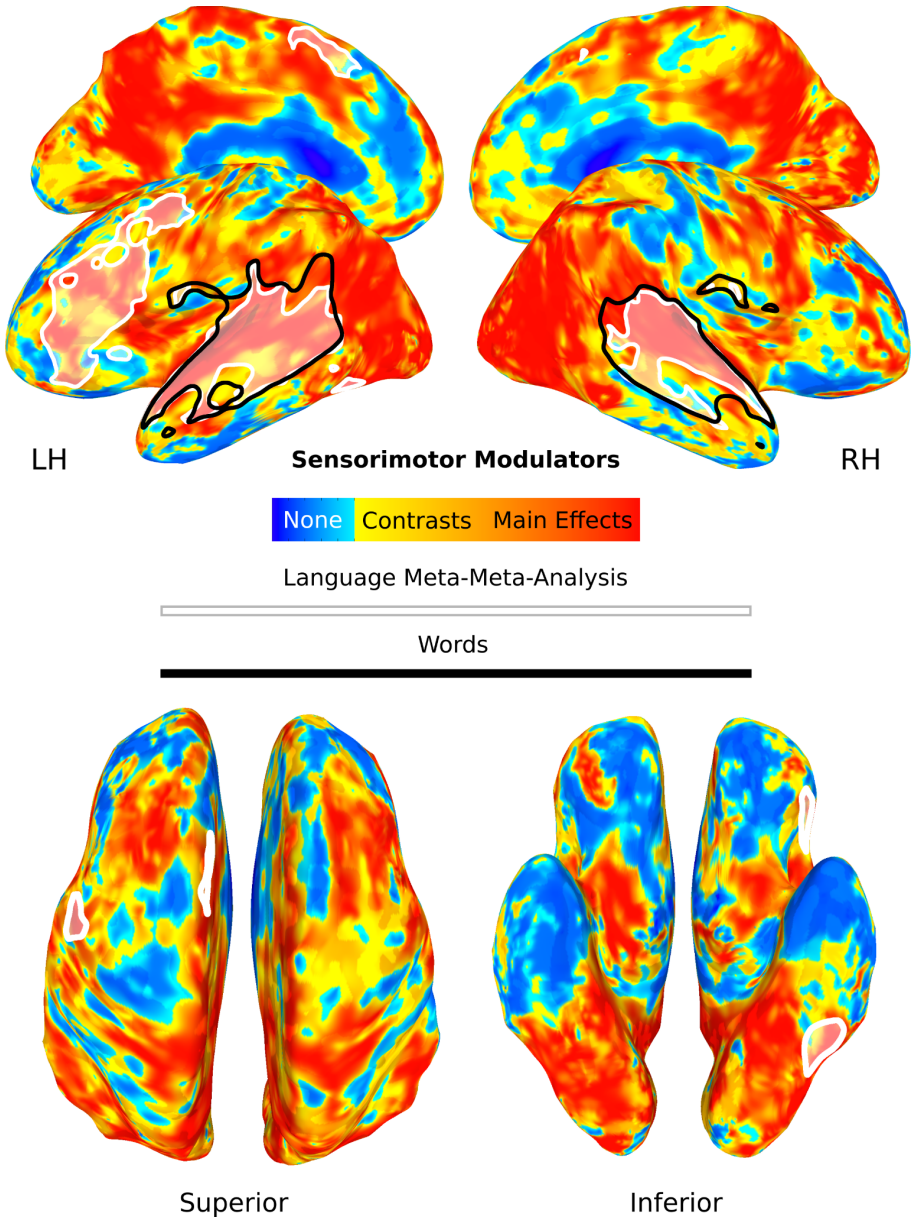

Figure 4. *Distribution of activity for words and their experiential associations. The black outline corresponds to the main effect of words and the language meta-meta-analysis (white outline) is included for comparison. The effect of the 11 sensorimotor modulators of words are coloured in red (e.g., foot/leg) and pairwise contrasts between modulators are coloured in yellow (e.g, foot/leg vs hand/arm). Insignificant voxels are in blues for comparison. All results were cluster-size corrected for multiple comparisons at* $\alpha = 0.01$ *and displayed with a minimal cluster of 20 voxels (540 ml).*

Hubs

In parallel with the overlap between the language meta-meta-analysis and the meta-analytic measure of centrality (Figure 3), we expected regions associated with words to be connectivity hubs. To test this hypothesis, we constructed individual voxel-wise networks in just grey matter using a

sliding-window approach and averaging over four measures of network centrality (degree, eigenvector, betweenness, and closeness). We clustered the centrality values using Ward’s minimum variance (66), which divided voxels in each time window into low- (given a value of 1) and high-centrality clusters (given a value of 2).

We first averaged across all windows and performed a linear mixed effects model for the group level analysis. The fixed effects were centrality (high and low), age, gender, and film with participant as a random effect. A correction for multiple comparisons using a cluster-size correction with multiple voxel-wise thresholds was applied (again, using $\alpha = 0.01$). Results show that high centrality voxels are significantly more active than low centrality voxels in most of the superior and middle temporal plane (Figure 5, colours) and that the main effect of words overlap with these regions (Figure 5, black outline). Indeed, 78.26% of the word voxels were high > low centrality (thresholded at $\alpha = 0.01$ with a minimum individual voxel p-value ≤ 0.001). Even when thresholding results further to include only the top 90% of the high centrality voxels, there was still a 39.00% overlap.

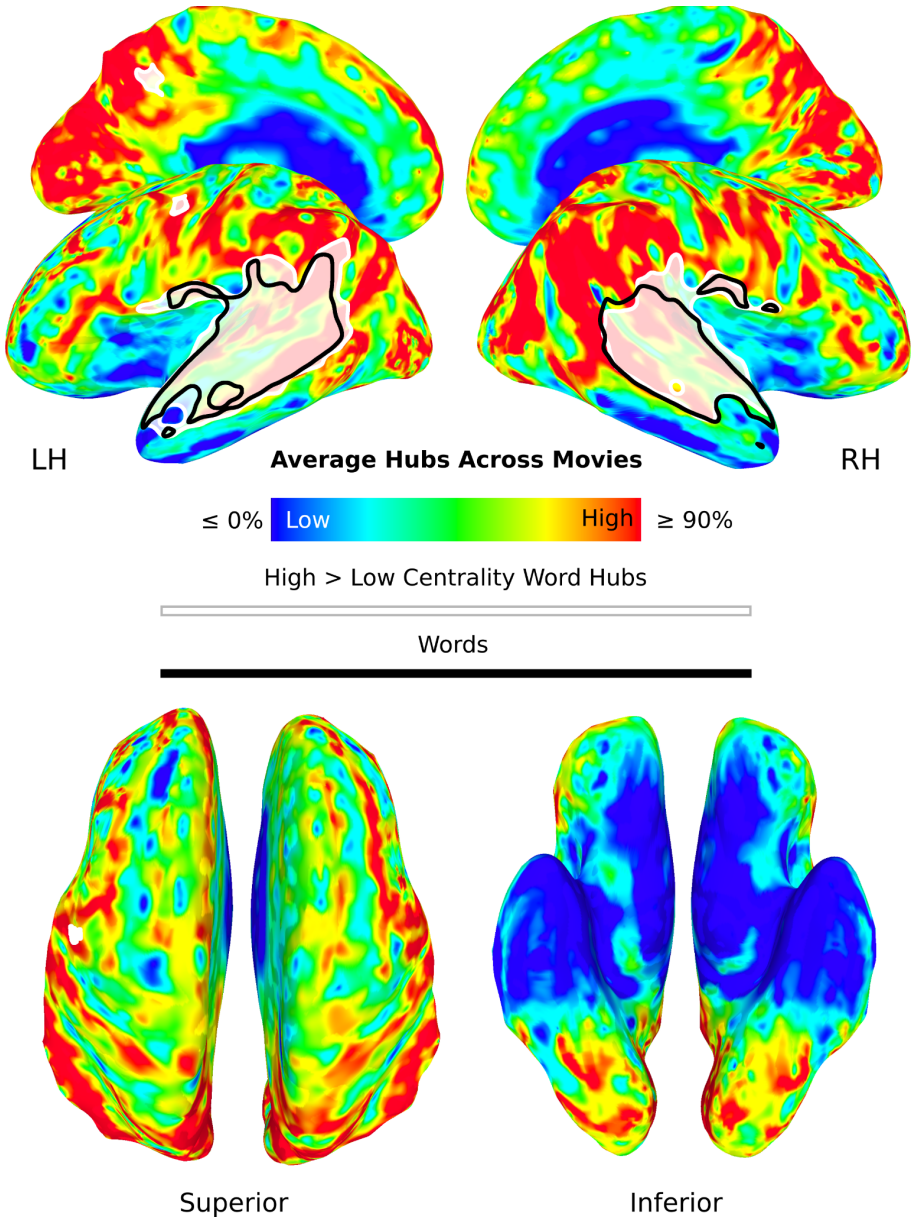

Figure 5. *Distribution of high versus low centrality regions. Dynamic functional connectivity was conducted and colours show high minus low centrality regions averaged over time. Significant high > low regions that also exceed a 90% value to be considered hubs are in red. Spatial independent component and dual regression analyses were also conducted on the high and low centrality timeseries across participants to find centrality hubs aggregated over time. Two components had a high correlation with the main effect of words (white shaded in outline) and were high > low centrality. The black outline corresponds to the main effect of words for comparison. All results were corrected for multiple comparisons.*

The linear mixed effects model collapses over films, not permitting separability of high centrality hubs into possibly distinct sets. Thus, we next attempted to formally determine whether regions delineated

by words formed a coherent set of hubs, independent of other sets in the aggregate across time. We accomplished this by performing group spatial independent component analysis (ICA) with 100 dimensions across participants’ dynamic high and low centrality timeseries. We then used dual regression to quantify the differences in high and low centrality, involving a direct contrast and thresholding using a correction for the 100 comparisons conducted. Finally, we computed the spatial correlation of each of the resulting contrasts with words, requiring a moderate or larger threshold (i.e., $r \geq 0.30$). Results reveal two contrasts correlated with words (IC 9, $r = 0.56$ and IC 15, $r = 0.53$). These were both significantly more associated with high than low centrality, and were combined for display purposes (Figure 5, shaded white outline). Collectively, these two high centrality hubs shared 80% of their voxels with words and were together spatially correlated with words at $r = 0.73$[5].

Thus, both the linear mixed effects and dual regression analyses suggest that voxels associated with words form a coherent set of connectivity hubs. However, we hypothesised that these only appear in the aggregate over time and do not exist as a whole on a moment-to-moment basis. That is, hubs are not fixed but are dynamic, e.g., moving around the superior and middle temporal lobes to coordinate distributed and varying peripheral regions. If this is the case, individual time windows should not be correlated with words as highly as in the ICA analysis. To examine this, we averaged across participants for each window for two of our films separately, ‘500 Days of Summer’ and ‘Citizenfour’. We thresholded the resulting timeseries voxels at 90% (i.e., ≥ 1.8) of the mean maximum value (i.e., 2) to isolate high centrality voxels. We then spatially correlated each thresholded time window with words. The mean correlation for ‘500 Days of Summer’ was $r = 0.03$ ($SD = 0.05$) with only one time window having a correlation ≥ 0.30 (and only 10.55% of windows with $r \geq 0.10$). The values for ‘Citizenfour’ were comparable, with the mean correlation being $r = 0.02$ ($SD = 0.04$) with no time windows having a correlation ≥ 0.30 (and only 6.08% of windows with $r \geq 0.10$). Performing ICA on these two films produced similar results as those in the prior paragraph, with two networks correlated with words $r \geq 0.30$ for each film (IC 17, $r = 0.58$ and IC 35, $r = 0.58$ for ‘500 Days of Summer’ and IC 31, $r = 0.49$ and IC 34, $r = 0.66$ for ‘Citizenfour’).

Collectively, these results suggest that the word hubs only appear in the aggregate over time and are present only in part on a moment-to-moment basis. To explore this further, we performed additional analyses to examine the connectivity between word hubs and the periphery at individual time windows. By the proposed model, when word voxels are acting as hubs, peripheral voxels should largely not be acting as hubs. To test this, we subtracted word voxels from the sensorimotor maps and vice versa (Figure 4, black outline and colours), creating word and periphery masks. We computed the

[5] Note that results need not have turned out this way, e.g., word regions could have been consistently low centrality regions.

mean cluster assignment for each window for each participant across voxels in each of these masks, again considering a mean of 90% (i.e., ≥ 1.8) as high centrality. The word mask acted as a hub by this metric on average of 21.54% (SD = 9.66%) of time windows across participants, lending support to analyses in the prior paragraph. In contrast, the peripheral mask had a value of ≥ 90% for only 0.07% (SD = 0.10%) of time windows. When the word mask was a hub, the peripheral mask was on average not a hub. When the word mask was a hub, the average hub variance was 0.12 compared to 0.22 for the peripheral mask.

Finally, it might be the case that highly central word hubs are not connected to the periphery, e.g., if the periphery is ‘extraneous’ and acting independently. To address this, we inspected the specific connectivity profiles of windows where the mean cluster assignment of the voxels from the word mask was ≥ 90% in ‘500 Days of Summer’. This analysis shows that when the word mask was a hub, it shared an average of 44% (SD = 2%) of connections with peripheral voxels across these 20 participants. This helps explain the low spatial correlation with the thresholded windows and words and the relatively low percentage of word masks acting as hubs at any given time window above. That is, when word voxels are hubs at any given moment, they are connected to a periphery, driving down correlations with word maps that have been aggregated over time and thus exclude the less central and more variable periphery.

Discussion

Summary

Classical and contemporary models of the neurobiology of language suggest that there are a small number of fixed ‘language regions’. Yet, this is juxtaposed with significant evidence that language processing is distributed throughout the entire brain. To reconcile these differences, we tested the hypothesis that ‘language regions’ result from using measures of central tendency and thresholding across heterogeneous language representations and processes. Such averaging minimises the activation in regions with greater spatial variance, while leaving activation in auditory input regions (in the case of vocal/heard languages) and connectivity hubs that coordinate those more variable peripheral regions.

Indeed, using both neuroimaging meta-analyses and film-fMRI, we show that language processing seems at first blush to occur in a very circumscribed set of fixed brain regions (Figures 1 and 4). A meta-analysis of 85 meta-analyses of different language representations and processes shows that the same superior and middle temporal and posterior inferior frontal regions are repeatedly activated, with little spatial variance (Figure 1). This same set of regions overlapped highly with those activated when

86 participants listened to words being spoken during one of 10 different films while undergoing fMRI (Figure 4, compare the black and white outlines).

This consistency in localisation disappears when words are no longer treated homogeneously (Figures 2 and 4). That is, meta-analyses of even the gross distinction between verbs and nouns suggest that large swaths of brain tissue participate in language processing (Figure 2). When making even finer distinctions in the film-fMRI data, the distributed nature of processing becomes even more salient (Figure 4, colours). Specifically, individual (e.g., auditory and visual) sensorimotor properties of words produce distributed patterns of activity, averaging about 4% of the brain per each of 11 sensorimotor dimensions beyond word processing regions (over which those properties are averaged). Words consist of multiple sensorimotor properties and their conjoined activity encompasses up to 67% of the remainder of the brain. This includes regions important for processing action, emotions, interoception, movement, somatosensation, and vision (among others).

Thus, averaging and thresholding lead to misleading minimisation of activity in regions sometimes regarded as ‘non language’ or ‘extraneous’ (compare Figures 1 and 2; Figure 4). We hypothesised that this leaves only auditory processing regions centred around primary auditory cortex and regions of high connectivity that coordinate those distant but no longer observable regions. Indeed, most ‘language regions’ are connectivity hubs as defined by degree centrality in a large-scale meta-analytic connectivity analysis (Figure 3) or the average of four different measures of centrality in the NNDb data (Figure 5). However, additional analysis shows that these connectivity hubs are not fixed. Rather, they are dynamic, with any individual hub occupying only some of what have become known as ‘language regions’ at any given moment and appearing together only when aggregating over time. On this moment-by-moment basis, language related hubs are connected to a periphery of less central and more variable brain regions that are typically averaged away.

Local vs Distributed

Collectively, these results beg for a reconsideration of localizationist accounts of the biological basis of language processing in the brain. Whether implicitly or overtly, it seems that the localizationist view of the 19th century, reconstituted in the imaging era of recent years, derives from confirmatory though potentially misleading results from thousands of neuroimaging studies. This is reflected in the widespread use of ‘language localisers’. These are ostensibly used for good reasons, to account for individual variation in the location of language regions (67–69) despite arguments against their usefulness (70–72). Regardless, a localiser task typically involves listening to intelligible language and to less intelligible language (among other variants), with analysis that requires averaging over many different kinds of linguistic representations and processes in both conditions, subtracting them, and

thresholding. It is therefore not surprising, based on the arguments advanced here, that across 45 languages, localisers simply show the same regions as seen in Figure 1 (73). The inappropriate adoption of the localisionist view is also reflected in contemporary model building in which language processing is said to be supported by only a small number of fixed regions (3, 5, 6, 74, 75). The implication of our results is that these models are necessarily incomplete, in that they do not account for all regions or that they are wrong in claiming that only ‘language regions’ process language.

There are several counter arguments that might be proffered in favour of maintaining more localizationist norms. These include that the more distributed results that we observe, particularly those outside of ‘language regions’, are 1) not representative of language processing but some other process like imagery or 2) limited to only ‘sensorimotor’ representations and ‘semantic’ or ‘conceptual’ processing. We next address these counter arguments, demonstrating that our distributed results are unlikely to be ‘non-linguistic’ and that they generalise to other language representations and processes. In the subsequent section, we further argue that a whole-brain model of the neurobiology of language is necessary for language comprehension and sketch a corresponding network architecture that might better account for the neurobiology of language than current models.

Doing Language?

A sceptic might argue that the ‘language regions’ we observe do in fact only process language and that the other more distributed regions are doing something possibly independent and nonlinguistic, like post-perceptual imagery, conceptual processing, or thought. The methods and stimuli that we used in the NNDb study when taken collectively help ensure that this is not the case. In the regression analysis, the sensorimotor properties were included as individual word amplitude modulators, meaning they are time locked to word processing activity and not likely other processes. The model also included auditory and visual nuisance regressors to control for other features from the films that might covary with those words. Moreover, the speech in films is continuous without pause between words, making it unlikely participants spent time imagining or conceptualising those words.

Further bolstering this argument, previous studies support the claim that those more distributed regions are doing something linguistic. Specifically, ‘language regions’ and other sensorimotor regions form networks during word processing and the latter are active within 50-150 ms after the word onset (27–30). Together, such results suggest that the sensorimotor component of the activation is part of and inseparable from the distributed representation of the words themselves (22). Our results support this view in that regions associated with words are connected to a large sensorimotor periphery on a moment-to-moment basis. This view is also consistent with results in other domains like vision and

memory showing that representations are not confined to but are maintained across reciprocally interconnected neurons and regions (76–78).

In contrast, the strongest claims involving ‘language localisers’ suggest that the putatively domain-general ‘multiple demand network’ (MDN) ‘shows no sensitivity to linguistic variables’ (79). This argument creates a false and sterile dichotomy (80). First, the MDN is often sensitive to ‘linguistic information’ albeit at reduced levels (8, 9, 79, 81–84). Second, analysis are done by averaging over large regions with different connectivity and cytoarchitecture profiles (85), comprising about 25% of the grey matter voxels in the brain[6]. Collectively, these regions are appropriately named as they have some of the highest measures of functional diversity in the human brain (86, 87). These functions are certainly not limited to domain-general processes (e.g., premotor cortices play clear roles in speech perception)(88). As functional diversity and ‘neural reuse’ (and ‘cortical recycling’) imply, these regions do many things and do contribute specifically, demonstrated when analyses do not average over individual language representations and processes (discussed next)(80, 89–91).

More Than Sensorimotor?

It is also unlikely that our argument that ‘language regions’ are the results from averaging is specific to only ‘sensorimotor’ representations and ‘semantic’ or ‘conceptual’ processing. That is, the entire set of distributed regions beyond the putative ‘language regions’ has been variously demonstrated to be involved in many other language representations and processes when they are probed. For example, people with global aphasia who are missing the entire ‘language network’ can still produce formulaic language (or multiword expressions), suggesting these linguistic representations are instantiated outside of ‘language regions’ (33). Despite this, there are less than five neuroimaging studies that we are aware of that distinguish formulaic language as a different type of linguistic representation. Yet, we have shown that novel sentences are processed in ‘language regions’ but when these same sentences are overlearned over a 15 day period, becoming formulaic, they are processed in the central sulcus and subcortical structures and not in ‘language regions’ (32).

Many other examples can be provided that non-‘language regions’ are involved in language beyond their role in sensorimotor representations and semantic or conceptual processing. To give a non-exhaustive overview, orbitofrontal cortex, among other regions, plays a role in indirect speech act comprehension (92). The dorsolateral prefrontal cortex (part of the putative MDN) plays numerous specific roles in ‘discourse management, integration of prosody, interpretation of nonliteral meanings, inference making, ambiguity resolution, and error repair’ (93). Dorsal medial prefrontal regions are

[6] Calculated using https://neurovault.org/images/57681/

involved in understanding speech acts, nonliteral meanings, and emotionally and socially laden language (94). These regions and the precuneus are involved in various aspects of generating and updating language based situation models (95, 96). More generally, these regions comprise aspects of the ‘default mode network’ which have been directly linked to ‘language regions’, inner speech, and language comprehension (97–101).

Other regions, like the anterior cingulate and parietal cortices implement necessary adaptive language control processes (102). The entire distributed motor system plays a number of roles in supporting speech perception (88). The insula and limbic structures like the amygdala are involved in affective prosody processing (103–105). Other subcortical structures like the basal ganglia (36), thalamus (106–108), and cerebellum (109) play numerous linguistic roles, including processing of syntax. Visual regions like the putative ‘visual word form area’ (110) and face and motion processing regions (111, 112) play roles in audio-only speech perception, even when there is no visual information available.

NOLB Model

Claiming that these various distributed language representations and processes are somehow non- or extra-linguistic or, worse, merely pragmatic is problematic. It threatens to reduce our real-world understanding of language to some core elements that in themselves cannot explain how the human brain manages to generate and extract information through language, e.g., to syntactic recursion (113, 114) or semantic compositionality (115). However, we suggest there is an even more fundamental reason for rejecting linguistic and nonlinguistic dichotomies for a whole-brain neurobiological model. That is, doing so is part of an account that helps resolve a fundamental problem in speech perception and language comprehension, namely, how the brain overcomes linguistic ambiguity. Next, we sketch a mechanistic model by which the brain uses both linguistic and nonlinguistic contextual information to predict language representations, constraining interpretation (12).

Language representations and their associated processes are ambiguous at all levels (for a review, see)(12). For example, there are no known acoustic features that unambiguously distinguish phonemes (called the ‘lack of invariance problem’). Words are not only homonymous (e.g., ‘bat’ and ‘bank’) but most are polysemous (e.g., the word ‘set’ has more than 450 meanings). Similarly, sentences and discourse can be syntactically and/or semantically ambiguous. How does the brain resolve all these ambiguities? We and others have taken proposals from Helmhotlz in vision (‘unconscious inference’, circa the 1860s) and Stevens and Halle in speech (‘analysis-by-synthesis’, circa the 1960s) and proposed that language ambiguity necessitates, in more modern parlance, prediction (12, 88, 116).

Specifically, the brain uses the ubiquitous amount of internal context (in the form of memories) and external or observable context it has available in the real world to make predictions about forthcoming language representations. To give an example, observable speech-associated mouth movements precede associated auditory information by about 100 ms. These can be used to make predictions about the acoustic patterns subsequently arriving in primary and surrounding auditory cortices, and constrain interpretation of which phonemes are intended. We have shown that this is in fact how audiovisual speech perception works, involving feedforward and feedback network interactions between visual, ventral pre- and primary motor cortices, and posterior superior temporal cortices (among others)(13, 15).

Because language is ambiguous and requires continual predictions derived from context, and context is always variable, the brain regions involved in speech perception and language comprehension will necessarily be highly variable. As these distributed regions are making predictions, they form connected networks. The only common denominator in these spatially variable and distributed networks are auditory input and connectivity hubs coordinating them. This has two implications relevant here: First, because, e.g., visual and motor regions are predicting intended phonemes, the distinction between ‘linguistic’ and ‘nonlinguistic’ is fuzzy at best. Second, averaging over the various representations and processes occurring in different and more variable distributed networks leaves only the aforementioned auditory and hub regions after thresholding.

What kind of network architecture can accommodate such a model? A classically modular architecture is not capable of achieving the flexibility needed to accommodate continually changing contexts and associated networks. An alternative architecture incorporates core-periphery structures that combine two network dynamics: The core is a collection of stable hubs that control a set of flexible regions, or periphery, that manifest remarkable temporal variability (117, 118). Core-periphery networks allow for high complexity, robustness to perturbations and rewiring of connections, and recovery following injury (118, 119). Our data appear to fit into a core-periphery framework with the identified ‘language regions’ corresponding to dynamic cores and the rest of the distributed activity (that is typically averaged away) constituting a dynamic periphery. It is left to future work to test this architecture with a whole-brain, voxel-wise core-periphery algorithm (120).

Implications

The studies and results here suggest that the historical neo-localizationist view of the neurobiology of language needs radical revision. By obscuring the distributed and whole-brain nature of language processing systems, static neuroimaging studies have inadvertently reinforced inferences made from lesion analyses in the 19th century. The latter suggest that gross language dysfunctions are caused by

damage to a small set of fixed regions[7]. Similarly, neuroimaging analyses based on averaging over linguistic 'apples and oranges' and thresholding, give the false impression that language processing occurs in a small set of fixed regions. Results here suggest that both the language problems caused by injury and the localizationist inferences that can be drawn from neuroimaging can be explained by considering that 'language regions' and 'the language network' are actually auditory input regions and connectivity hubs. This implies that aphasia is the result of damaging the core regions that coordinate a whole-brain distribution of regions that process language and not the result of damaging a putative 'language network'. Indeed, empirical evidence demonstrates that damage to connectivity hubs is more likely to underlie various disorders, including aphasia (122–127). This implies that a different (perhaps core-periphery) model of the neurobiology of language is needed to replace contemporary models to help advance treatment for aphasia (where effect sizes are low on average)(128). This might involve focusing therapy around individualised and preserved linguistic representations and processes in the periphery.

More generally, it is perhaps obvious that our conclusions about averaging and thresholding can be extended to every domain in which psychological ontologies are probed with neuroimaging data (42, 129). It is hard to think of any neuroimaging study in which stimuli do not contain multiple representations or processes that are mathematically integrated before thresholding. Therefore, what is revealed by most studies and neuroimaging meta-analyses is more likely to reflect underlying connectivity hubs rather than the full distribution of regions involved. This conclusion probably scales with the complexity of the behaviour under investigation. For example, visual object processing studies average over many different stimuli whose sensorimotor properties and affordances greatly vary, leaving only core regions involved in complex visual processing (like the fusiform gyrus). In contrast, language and social processing are at a pinnacle of human functioning and likely require more dynamic hubs orchestrating far more regions given the complexity of these processes. This implies that we need much more consideration of the representations and processes engaged by the stimuli and tasks we give participants and considerably more emphasis on individual differences (129–132).

[7] Though William James might have shared our view. As quoted in the title of the present article, he said: '*There is no "centre of Speech" in the brain any more than there is a faculty of Speech in the mind. The entire brain, more or less, is at work in a man who uses language.*' (121)

Methods and Materials

Meta-Analyses

Averaging

To conduct the neuroimaging ‘meta-meta-analysis’, we manually searched the Neurosynth (133) and BrainMap (134) databases (accessed in August 2022) for available terms and studies related to language representations (e.g., ‘phonological’, ‘words’, ‘sentences’) and associated processes (e.g., ‘speech’, ‘semantics’, ‘syntax’). We excluded terms and study descriptions pertaining to music and reading or having an obvious visual element. This resulted in 57 Neurosynth terms and 28 BrainMap searches that were used to assemble the meta-meta-analysis (Table 1).

Specifically, the Neurosynth meta-analyses are already generated and we simply downloaded then from neurosynth.org in MNI space with a false discovery rate corrected threshold of $p \leq 0.01$ (from dataset version 0.7, released July, 2018)(133)[8]. These had been constructed by extracting 507,891 activation coordinates reported in 14,371 neuroimaging articles and collating a list of 1,334 high frequency terms that occur in more than 20 of those articles. Coordinates for each high frequency language related term are aggregated and compared to the coordinates reported for articles without those terms (Table 1, left). For BrainMap meta-analyses, meta-data from the ‘Sleuth’ application (version 3.0.4) was used to search for papers (Table 1, right) and corresponding coordinates, and the ‘GingerALE’ application (version 3.0.2) was used to perform activation likelihood meta-analyses on those coordinates (134–137)[9]. These were done in MNI space and thresholded using false discovery rate correction of $p \leq 0.01$ to match meta-analyses done in Neurosynth.

Each of the resulting 85 meta-analyses were further preprocessed to have a minimum cluster size of 50 voxels (400 ml). The spatial pattern of activity for the combined Neurosynth meta-analyses were highly correlated with the BrainMap meta-analyses ($r = 0.76$). Thus, the results were combined into one neuroimage by count using ‘3dmerge’ available in the AFNI software package (138). Additionally, we conducted a one-sample ‘group’ level statistical analysis with ‘software package’ as a covariate (i.e., Neurosynth or BrainMap) and a minimum of five contributing meta-analyses using ‘3dttest++’, also available in AFNI (138). Results were thresholded using a false discovery rate correction of $p \leq 0.01$ and a minimum cluster size of 50 voxels (400 ml).

[8] See https://neurosynth.org/ and https://github.com/neurosynth/neurosynth-data

[9] See https://brainmap.org/

Table 1. *Terms and searches used to conduct the meta-meta-analysis (Figure 1).*

Meta-Analyses	
Neurosynth Terms (N = 57)	**BrainMap Searches (N = 28)**
acoustic, aphasia, articulatory, audiovisual, auditory, auditory, cortex, auditory, stimuli, broca, communication, comprehension, english, fluency, hearing, language, language, comprehension, language, network, languages, lexical, lexical, decision, linguistic, listened, listening, names, naming, noise, oral, phonological, primary, auditory, pseudowords, semantically, semantic, semantic, memory, semantics, sentence, sentence, comprehension, sentences, sound, sounds, speaker, speakers, speaking, speech, speech, perception, speech, production, symbolic, syntactic, tone, verbal, verbal, fluency, verbal, working, verb, verbs, voice, word, word, pairs, word, recognition, words	experiments beh dom is cog language phonology, language speech, language syntax, language experiments beh dom is perception audition experiments context is language experiments stimulus type is reversed speech, sentences, stories, syllables paradigms affective words, divided auditory attention, figurative language, freelist word recall, lexical decision, naming covert, naming overt, passive listening, pitch monitor discrimination, recitation and repetition covert, recitation and repetition overt, semantic monitor discrimination, syntactic discrimination, tone monitor discrimination, word generation covert, word generation overt, word imageability, word stem completion overt

Distributed

To determine the extent of language processing in the brain when not averaging over all representations, the NeuroQuery database was queried for ‘verbs’ and ‘nouns’ as an example of differentiable representations that are often averaged over, e.g., as would necessarily be the case in the ‘sentences’ meta-analysis (accessed, August 2022)(139)[10]. Unlike Neurosynth, NeuroQuery produces meta-analyses that are predictions of where in the brain a study pertaining to the input terms are likely to report activity. We used this approach over Neurosynth because it is able to make use of coordinates from nearly 10 times more articles which would presumably make results more robust for relatively less frequently studied topics. In particular, NeuroQuery produced a meta-analysis from 662 articles for the term ‘verbs’ (compared to 107 in Neurosynth) and 889 for ‘nouns’ (compared to 100 in Neurosynth). Results are presented unthresholded in Figure 2. To determine significant regions of activations reported in the Results section, the meta-analyses were subtracted from each other and thresholded at $z = 3.89$ ($p \leq 0.0001$), excluding any overlap and using a minimum cluster size of 50

[10] https://neuroquery.org/

vowels (400 ml). The patterns of activity for these thresholded results were similar to but more robust than the $p \leq 0.01$ false discovery rate correct ‘verbs’ and ‘nouns’ meta-analyses from Neurosynth.

Hubs

We developed a new meta-analytic network connectivity hub metric to examine whether the resulting ‘language regions’ from the meta-meta-analysis are in fact hubs (64, 65). First, we used the Neurosynth software package to conduct 165,953 voxel-wise co-activation meta-analyses across 14,371 studies. Similar to the logic of functional connectivity, the regular co-activation of two or more regions is suggestive that those regions form functional connections or a network. Each meta-analysis was thresholded at $p \leq 0.01$ false discovery rate corrected and combined by count using ‘3dmerge’ available in the AFNI software package (138). This corresponds to the number of connections (edges) that a voxel (node) has and, thus, is a meta-analytic equivalent of degree centrality. The resulting map was spatially converted to z-scores using a whitematter mask generated by Freesurfer to limit results to grey matter (140, 141). This is presented unthresholded in Figure 3. Hubs were defined as voxels with 3.89 standard deviations more connections than the mean (equivalent to $p \leq 0.0001$). Overlaps with meta-meta-analysis results were tested at this z-value as well as $z = 3.29$ ($p \leq 0.001$), $z = 2.58$ ($p \leq 0.01$), and $z = 1.96$ ($p \leq 0.05$).

NNDb

Neuroimaging

Our fMRI data derives from 86 participants who watched one of 10 previously unseen full-length films from 10 different genres (42 females, age range of 18–58 years, M = 26.81, SD = 10.09 years). All participants were right-handed, native English speakers with normal hearing and vision (corrected), without claustrophobia or neuro-psychiatric illness, and not taking medication, all determined by self-report. The study was approved by the ethics committee of University College London and participants provided written informed consent to take part in the study and share their anonymised data.

Functional MRI scans were collected on an 1.5T Siemens MAGNETOM Avanto using a 32-channel head coil and a multiband EPI sequence (TR = one second, TE = 54.8 ms, flip angle of 75°, 40 interleaved slices, resolution = 3.2 mm isotropic, multiband factor = four, no in-plane acceleration), with breaks determined by participants about every 40 minutes. The films were one hour and 56 minutes long on average, meaning there was a M = 6,998.20 TRs (SD = 1068.92; Range is 5,470 to 8,882 TRs). A 10 minute T1-weighted anatomical scan was collected after the functional scans using a MPRAGE sequence (TR = 2.73 seconds, TE = 3.57 ms, 176 sagittal slices, resolution = 1.0 mm).

Unless otherwise noted, all preprocessing and data visualisation was done with the AFNI software package and specific programs are indicated where appropriate (138)[11]. First, anatomical images were corrected for intensity non-uniformity, deskulled, and nonlinearly aligned to an MNI template (i.e., MNI152_2009_template_SSW.nii.gz; using ‘@SSwarper’). Next, Freesurfer’s ‘recon-all’ was run with default parameters (version 7.0)(140, 141)[12]. This allowed us to segment anatomical images into ventricle and white matter regions of interest to be used to make nuisance regressors.

Preprocessing of the functional data was done using ‘afni_proc.py’[13] and timeseries are available as derivatives on OpenNeuro[14]. Blocks were included for despiking (‘despike’), slice-timing correction (‘tshift’), volume registration to the run with the least number of outliers and warping to MNI space (in one step, ‘align’, ‘tlrc’, and ‘volreg’), masking (‘mask’), six mm blurring (‘blur’), scaling to a mean of 100 (‘scale’), and regression (‘regress’). The latter used ‘3dDeconvolve’ to remove demeaned motion and motion derivatives, a polynomial degree of two (‘polort 2’), bandpass filtering (0.01 to 1), ventricular activity (first three principle components), and white matter activity from the timeseries on a per run basis, in one step. The resulting timeseries were then edited to account for minor temporal delays caused by film pauses (on the order of milliseconds). Finally, hand labelled independent component based noise regressors were removed from these timeseries (for details, see)(142).

Averaging

Regression

We used the AFNI program ‘3dDeconvolve’ to conduct a single duration and amplitude modulated multiple linear regression model on the preprocessed timeseries to determine the distribution of brain activity for words and their associated sensorimotor properties. This approach allows us to find a ‘main effect’ for words as well as determine how the sensorimotor properties of those words modulate the amplitude of their response. To do this, films were first annotated for word onset and duration using a combination of machine learning based speech-to-text translation, dynamic time warping to the subtitles, and manual correction (see)(142). Next, we found the intersection of each word annotation (M = 10,323.80, SD = 2,909.40 words) with the 39,707 words in the Lancaster database of sensorimotor norms (143). The norms themselves were created by having participants rate words on 11 experiential dimensions, i.e., auditory, foot/leg, gustatory, hand/arm, haptic, head, interoceptive, mouth, olfactory, torso, and visual (143). On average, 95.8% of the words in the films had a

[11] https://afni.nimh.nih.gov/

[12] http://www.freesurfer.net/

[13] https://afni.nimh.nih.gov/pub/dist/doc/program_help/afni_proc.py.html

[14] https://openneuro.org/datasets/ds002837/versions/2.0.0

corresponding entry in the database. These were included with the words as 11 sensorimotor word modulators.

We also include three additional ‘nuisance’ modulators to control for the confounding effects of auditory and visual stimulation and word frequency (as used in prior studies)(83, 144–146). Specifically, to account for auditory and visual information that might be correlated with sensorimotor norms in the film, two ‘low-level’ values were created for each of the words having norms. Sound energy was measured as the root-mean square acoustic energy of the film audio track and was calculated every 100 ms using the Python library ‘librosa’ (147). Contrast luminance, which measures the standard deviation in luma (brightness) values of the pixels in an image, was computed at every frame in the film using the Python library ‘OpenCV’ (148).

Both the sound energy and contrast luminance values were averaged over the duration of the words. For instance, if a word had a duration of 200 ms, two values of sound energy (each at 100 ms) would be averaged and 5000 values of contrast luminance would be averaged together (each at 0.04 ms). In the event that the word duration was smaller than the sampling rate of either control regressor, the value of one sampling step was assigned to the word. Thus, for instance, if a word had a duration of 90 ms, the sound energy value assigned to the word would be from one 100 ms value, and the contrast luminance would be the average of 2250 values. Across words the mean sound energy was $8x10^{-3}$ (SD = $4.5x10^{-3}$) and the mean contrast luminance was 53.7 (SD = 17.6). Finally, to account for word frequency possibly correlating with specific sensorimotor norms, we created a third ‘nuisance’ modulator value. Because our participants all lived in the United Kingdom, we used the log-transform of word frequency database ‘Subtlex UK’ (149). Only one word did not have an associated word frequency and the mean word frequency across words was 6.2 (SD = 1.2).

In the regression, a block function was used to model the hemodynamic response function whose shape varied by duration and 14 amplitude modulators. These were the 11 sensorimotor norms, contrast luminance, sound energy, and word frequency. Words without sensorimotor norms were modelled with only a duration modulated block function as a separate regressor (M = 4.2%, SD = 0.8% of words in films). Finally, we included a duration modulated regressor for film times when no words were present (M = 1707, SD = 179 seconds). Regression for words with sensorimotor norms had the following form:

*word onset * auditory, foot/leg, ... visual, luminance, soundpower, frequency : duration*

LME

The resulting words and 11 sensorimotor amplitude modulated beta coefficient maps from the multiple linear regression were input into a linear mixed effects model using ‘3dLME’ (150). Age, gender, and film watched for each participant were included as covariates with age centred around the mean. We set participant as a random effect, whereby the intercept of the slope was allowed to vary by a small random amount compared to the group average for each participant. We computed the ‘baseline’ contrasts for words, all 11 sensorimotor norms, and pairwise contrasts between norms.

The results of these contrasts were corrected for multiple comparisons using a multi-thresholding approach (modelled after)(151). First, we estimated the smoothness and autocorrelation function of neighbouring voxels using the ‘3dFWHMx’ command in AFNI (152). Then we ran ‘3dClustSim’ over six uncorrected individual voxel p-values (0.05, 0.02, 0.01, 0.005, 0.002, and 0.001) to achieve an alpha (α) threshold of 0.01. Using the significant cluster sizes whereby faces or edges need to touch and voxels are contiguous if they are either positive or negative at each p-threshold, we merged the thresholded maps at each p-threshold to obtain significant voxels ($\alpha = 0.01$). We discuss and display all results using a minimum cluster size of 20 voxels (540 ml).

Hubs

We next sought to determine if the voxels corresponding to the main effect of words in the 3dLME model correspond to hubs. We did this using measures of centrality, how important a node is for the integrity and information-flow within a network. Centrality can be determined using various metrics that provide different information on the role of the node of interest in the network. We measured degree, eigenvector, closeness, and betweenness centrality in an effort to assure that we are robustly characterising centrality across metrics (44, 153–155).

Degree centrality is the sum of inward and outward connections from a node. Eigenvector centrality is a measure of influence on a network, meaning that a high-connectivity node linked to nodes of high connectivity will have higher eigenvector centrality (i.e., be more influential) than a high-connectivity node linked to low-connectivity nodes. Betweenness centrality measures the shortest paths that pass through a given node and closeness centrality measures the inverse of the distance of shortest paths passing through the node. Although these centrality metrics provide different details on the importance of a node, they are correlated to one-another (155, 156).

We calculated these four measures of centrality on a voxel-wise basis and in a dynamic manner. To do this, we constructed time-varying connectivity matrices using a sliding-window approach. First, the timeseries was resampled from three to a four mm^3 resolution (using ‘3dresample’) because the

computation time needed to conduct analyses at the native three mm resolution was estimated to take too long. Resampled timeseries were then divided into windows of one minute length, sliding every 10 seconds to allow for a 50 second overlap between one window and the subsequent window. Next, a pairwise Pearson’s product moment correlation coefficient was computed on a voxel-wise basis at each window using the AFNI program ‘3dDegreeCentrality’ (152). Finally, the resulting correlation matrices were proportionally thresholded to obtain a 10% sparsity in each time window. These values were considered a connection between two voxels and used to build connectivity matrices which were used to calculate the centrality values described.

Despite differences, these four centrality metrics had a significant Spearman’s ranking correlation with one another at the group level (i.e., across time windows and across participants; $M_\rho = 0.94$, $SD_\rho = 0.02$; $p \leq 0.001$). Thus, to simplify analyses and conform to our hubs/periphery hypotheses, we clustered the four centrality values using Ward’s minimum variance (66), which consistently divided voxels in each time window into two groups ($M = 2.00$, $SD = 0.02$) across all participants, one high-centrality (labelled 1) and one low-centrality cluster (labelled 2). In rare cases, the clustering method detected greater than two clusters of centrality, but since the vast majority of windows divided centrality values into two groups, we recomputed the few outlier time windows by forcing them to split the data into two clusters to investigate spatial variations in connectivity over time.

Next, we divided the resulting windowed centrality timeseries into a high and a low centrality timeseries for each participant and averaged across all windows. The resulting maps were used to conduct a linear mixed effects model with ‘3dLME’ (described above)(150). The fixed effects were centrality (high and low), age, gender, and film with participant as a random effect. A correction for multiple comparisons using a cluster-size correction with multiple voxel-wise thresholds was applied (again, as described above, using $\alpha = 0.01$).

We then submitted the full windowed high and low centrality timeseries from all participants to a group spatial independent component analysis (ICA) using the default parameters of ‘MELODIC’ (version 3.15), reducing to a 100-dimensional subspace (157)[15]. These results were then used in a dual regression to estimate a version of the 100 resulting components for each participant so group statistics could be calculated. Specifically, this approach first regresses the group-spatial-maps into each participant’s timeseries to give a set of timecourses. These timecourses are then regressed onto the timeseries to get 100 participant-specific spatial maps for each high and low centrality. These were then directly contrasted using a t-test and thresholded using voxel-wise z-value equivalent to $p \leq$

[15] www.fmrib.ox.ac.uk/fsl

0.0001 (i.e., 0.01 divided by the number of components) and a minimum cluster threshold of 20 voxels (540 ml) to protect against multiple comparison issues. Finally, we computed the spatial correlation of each of these contrasts with words from the above 3dLME analysis to locate a hub or hubs corresponding to words.

We also averaged each time window across participants for ‘500 Days of Summer’ and ‘Citizenfour’ separately to conduct two additional MELODIC ICAs, again with 100 dimensions, with resulting components thresholded using a mixture modelling approach (157). All other analyses are as described in the Results section.

Acknowledgments

This work was partially supported by a doctoral training grant given to SA from the UKRI Biotechnology and Biological Sciences Research Council (BBSRC) as part of the London Interdisciplinary Biosciences Consortium (LIDo). We would like to thank the Birkbeck-UCL Centre for Neuroimaging (BUCNI) for supporting the NNDb. We also thank Annette Glotfelty for helpful comments on the manuscript and The LAB Lab more generally for continued discussion. JIS would also like to thank the patience of a special Banana.

Author Contributions

JIS and SA conceived of the study. JIS did all fMRI preprocessing. JIS did all meta-analyses. SA developed and did analyses herein with ‘500 Days of Summer’. SA and BW extended this initial analyses to all 10 NNDb films. SA wrote the first draft of the manuscript. JIS wrote the second version of the manuscript, incorporating meta-analyses. SLS helped with writing the final version of the manuscript. JIS made all of the figures.

Conflicting Interests

The authors declare that they have no conflicts of interest.

Open Science Statement

This study was not preregistered. A version of this work is available in SA’s Doctoral thesis, ‘A Model of the Network Architecture of the Brain that Supports Natural Language Processing’. A version of this paper is available as a preprint on bioRxiv. The preprocessed data used in analyses is publically available on OpenNeuro (https://openneuro.org/datasets/ds002837/versions/2.0.0). Code used to generate results will be made available on GitHub (https://github.com/lab-lab/).

References

1. G. K. Y. Iii, Localization of Language Function in the Twentieth Century. *J. Hist. Neurosci.* **18**, 283–290 (2009).

2. J. E. Bogen, G. M. Bogen, Wernicke’s region--Where is it? *Ann. N. Y. Acad. Sci.* **280**, 834–843 (1976).

3. N. Geschwind, The organization of language and the brain. *Science* **170**, 940–944 (1970).

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

5. G. Hickok, D. Poeppel, The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).

6. J. P. Rauschecker, S. K. Scott, Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12**, 718–724 (2009).

7. A. D. Friederici, The brain basis of language processing: from structure to function. *Physiol. Rev.* **91**, 1357–1392 (2011).

8. E. Diachek, I. Blank, M. Siegelman, J. Affourtit, E. Fedorenko, The Domain-General Multiple Demand (MD) Network Does Not Support Core Aspects of Language Comprehension: A Large-Scale fMRI Investigation. *J. Neurosci.* **40**, 4536–4550 (2020).

9. I. A. Blank, E. Fedorenko, Domain-General Brain Regions Do Not Track Linguistic Input as Closely as Language-Selective Regions. *J. Neurosci.* **37**, 9999–10011 (2017).

10. K. L. Campbell, L. K. Tyler, Language-related domain-specific and domain-general systems in the human brain. *Curr Opin Behav Sci* **21**, 132–137 (2018).

11. E. Fedorenko, The role of domain-general cognitive control in language comprehension. *Front. Psychol.* **5**, 335 (2014).

12. J. I. Skipper, “The NOLB model: A model of the natural organization of language and the brain” in *Cognitive Neuroscience of Natural Language Use , (pp*, R. M. Willems, Ed. (Cambridge University Press, xiv, 2015), pp. 101–134.

13. J. I. Skipper, V. van Wassenhove, H. C. Nusbaum, S. L. Small, Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb. Cortex* **17**, 2387–2399 (2007).

14. J. I. Skipper, S. Goldin-Meadow, H. C. Nusbaum, S. L. Small, Gestures orchestrate brain networks for language understanding. *Curr. Biol.* **19**, 661–667 (2009).

15. J. I. Skipper, H. C. Nusbaum, S. L. Small, Listening to talking faces: motor cortical activation during speech perception. *Neuroimage* **25**, 76–89 (2005).

16. J. I. Skipper, S. Goldin-Meadow, H. C. Nusbaum, S. L. Small, Speech-associated gestures, Broca’s area, and the human mirror system. *Brain Lang.* **101**, 260–277 (2007).

17. J. M. Fuster, Distributed memory for both short and long term. *Neurobiol. Learn. Mem.* **70**,

268–274 (1998).

18. J. Rissman, A. D. Wagner, Distributed representations in memory: insights from functional brain imaging. *Annu. Rev. Psychol.* **63**, 101–128 (2012).

19. T. B. Christophel, P. C. Klink, B. Spitzer, P. R. Roelfsema, J.-D. Haynes, The Distributed Nature of Working Memory. *Trends Cogn. Sci.* **21**, 111–124 (2017).

20. J. I. Skipper, Echoes of the spoken past: how auditory cortex hears context during speech perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130297 (2014).

21. E. Yee, S. L. Thompson-Schill, Putting concepts into context. *Psychon. Bull. Rev.* **23**, 1015–1027 (2016).

22. F. Pulvermüller, Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* **6**, 576–582 (2005).

23. G. Vigliocco, D. P. Vinson, J. Druks, H. Barber, S. F. Cappa, Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neurosci. Biobehav. Rev.* **35**, 407–426 (2011).

24. J. R. Binder, R. H. Desai, The neurobiology of semantic memory. *Trends Cogn. Sci.* **15**, 527–536 (2011).

25. F. Carota, R. Moseley, F. Pulvermüller, Body-part-specific representations of semantic noun categories. *J. Cogn. Neurosci.* **24**, 1492–1509 (2012).

26. A. G. Huth, W. A. de Heer, T. L. Griffiths, F. E. Theunissen, J. L. Gallant, Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).

27. M. Kiefer, E.-J. Sim, B. Herrnberger, J. Grothe, K. Hoenig, The Sound of Concepts: Four Markers for a Link between Auditory and Conceptual Brain Systems. *Journal of Neuroscience* **28**, 12224–12230 (2008).

28. A. M. García, *et al.*, How meaning unfolds in neural time: Embodied reactivations can precede multimodal semantic effects during language processing. *Neuroimage* **197**, 439–449 (2019).

29. Y. Shtyrov, A. Butorina, A. Nikolaeva, T. Stroganova, Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E1918–23 (2014).

30. L. J. MacGregor, F. Pulvermüller, M. van Casteren, Y. Shtyrov, Ultra-rapid access to words in the brain. *Nat. Commun.* **3**, 711 (2012).

31. F. M. M. Citron, Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain Lang.* **122**, 211–226 (2012).

32. J. I. Skipper, *et al.*, Reorganization of the Neurobiology of Language After Sentence Overlearning. *Cereb. Cortex* (2021) https:/doi.org/10.1093/cercor/bhab354.

33. D. Van Lancker Sidtis, J. J. Sidtis, Cortical-subcortical production of formulaic language: A review of linguistic, brain disorder, and functional imaging studies leading to a production model. *Brain Cogn.* **126**, 53–64 (2018).

34. E. Fedorenko, A. Nieto-Castañon, N. Kanwisher, Lexical and syntactic representations in the brain: an fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia* **50**, 499–513 (2012).

35. A. Lopopolo, A. van den Bosch, K.-M. Petersson, R. M. Willems, Distinguishing syntactic operations in the brain: Dependency and phrase-structure parsing. *Neurobiology of Language* **2**, 152–175 (2021).

36. S. A. Kotz, M. Schwartze, M. Schmidt-Kassow, Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. *Cortex* **45**, 982–990 (2009).

37. M. Heard, Y. S. Lee, Shared neural resources of rhythm and syntax: An ALE meta-analysis. *Neuropsychologia* **137**, 107284 (2020).

38. R. M. Willems, P. Hagoort, D. Casasanto, Body-specific representations of action verbs: neural evidence from right- and left-handers. *Psychol. Sci.* **21**, 67–74 (2010).

39. I. M. Lyons, *et al.*, The role of personal experience in the neural processing of action-related language. *Brain Lang.* **112**, 214–222 (2010).

40. S. L. Beilock, I. M. Lyons, A. Mattarella-Micke, H. C. Nusbaum, S. L. Small, Sports experience changes the neural processing of action language. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 13269–13273 (2008).

41. M. W. Burton, D. C. Noll, S. L. Small, The anatomy of auditory word processing: individual variability. *Brain Lang.* **77**, 119–131 (2001).

42. C. Hawco, E. W. Dickie, G. Jacobs, Z. J. Daskalakis, A. N. Voineskos, Moving beyond the mean: Subgroups and dimensions of brain activity and cognitive performance across domains. *Neuroimage* **231**, 117823 (2021).

43. X. Liao, A. V. Vasilakos, Y. He, Small-world human brain networks: Perspectives and challenges. *Neurosci. Biobehav. Rev.* **77**, 286–300 (2017).

44. M. P. van den Heuvel, O. Sporns, Network hubs in the human brain. *Trends Cogn. Sci.* **17**, 683–696 (2013).

45. B. S. Khundrakpam, *et al.*, Developmental changes in organization of structural brain networks. *Cereb. Cortex* **23**, 2072–2085 (2013).

46. K. Wu, *et al.*, Age-related changes in topological organization of structural brain networks in healthy individuals. *Hum. Brain Mapp.* **33**, 552–568 (2012).

47. E. H. J. Nijhuis, A.-M. van Cappellen van Walsum, D. G. Norris, Topographic hub maps of the human structural neocortical network. *PLoS One* **8**, e65511 (2013).

48. E. Bagarinao, *et al.*, Identifying the brain’s connector hubs at the voxel level using functional connectivity overlap ratio. *Neuroimage* **222**, 117241 (2020).

49. M. A. Bertolero, B. T. T. Yeo, M. D’Esposito, The modular and integrative functional architecture of the human brain. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6798–807 (2015).

50. E. M. Gordon, *et al.*, Three Distinct Sets of Connector Hubs Integrate Human Brain Function.

Cell Rep. **24**, 1687–1695.e4 (2018).

51. K. Hwang, M. A. Bertolero, W. B. Liu, M. D’Esposito, The Human Thalamus Is an Integrative Hub for Functional Brain Networks. *J. Neurosci.* **37**, 5594–5607 (2017).

52. K. Hilger, M. Ekman, C. J. Fiebach, U. Basten, Intelligence is associated with the modular structure of intrinsic brain networks. *Sci. Rep.* **7**, 16088 (2017).

53. J. Li, *et al.*, High transition frequencies of dynamic functional connectivity states in the creative brain. *Sci. Rep.* **7**, 46072 (2017).

54. A. S. Mandal, R. Romero-Garcia, M. G. Hart, J. Suckling, Genetic, cellular, and connectomic characterization of the brain regions commonly plagued by glioma. *Brain* **143**, 3294–3307 (2020).

55. D. Meunier, S. Achard, A. Morcom, E. Bullmore, Age-related changes in modular organization of human brain functional networks. *Neuroimage* **44**, 715–723 (2009).

56. J. D. Power, B. L. Schlaggar, C. N. Lessov-Schlaggar, S. E. Petersen, Evidence for hubs in human functional brain networks. *Neuron* **79**, 798–813 (2013).

57. X. Wen, *et al.*, First-year development of modules and hubs in infant brain functional networks. *Neuroimage* **185**, 222–235 (2019).

58. T. Bolt, J. S. Nomi, M. Rubinov, L. Q. Uddin, Correspondence between evoked and intrinsic functional brain network configurations. *Hum. Brain Mapp.* **38**, 1992–2007 (2017).

59. C. Fauchon, *et al.*, The Modular Organization of Pain Brain Networks: An fMRI Graph Analysis Informed by Intracranial EEG. *Cereb Cortex Commun* **1**, tgaa088 (2020).

60. S. Fuertinger, B. Horwitz, K. Simonyan, The Functional Connectome of Speech Control. *PLoS Biol.* **13**, e1002209 (2015).

61. F. de Pasquale, S. Della Penna, O. Sporns, G. L. Romani, M. Corbetta, A Dynamic Core Network and Global Efficiency in the Resting Human Brain. *Cereb. Cortex* (2015) https:/doi.org/10.1093/cercor/bhv185.

62. W. R. Uttal, *Reliability in Cognitive Neuroscience: A Meta-Meta-Analysis* (MIT Press, 2013).

63. S. Aliko, J. Huang, F. Gheorghiu, S. Meliss, J. I. Skipper, A naturalistic neuroimaging database for understanding the brain using ecological stimuli. *Sci Data* **7**, 347 (2020).

64. N. A. Crossley, *et al.*, Cognitive relevance of the community structure of the human brain functional coactivation network. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11583–11588 (2013).

65. N. A. Crossley, P. T. Fox, E. T. Bullmore, Meta-connectomics: human brain network and connectivity meta-analyses. *Psychol. Med.* **46**, 897–907 (2016).

66. J. H. Ward, Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).

67. E. Fedorenko, P.-J. Hsieh, A. Nieto-Castañón, S. Whitfield-Gabrieli, N. Kanwisher, New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J.*

Neurophysiol. **104**, 1177–1194 (2010).

68. A. Nieto-Castañón, E. Fedorenko, Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage* **63**, 1646–1669 (2012).

69. T. L. Scott, J. Gallée, E. Fedorenko, A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cogn. Neurosci.* **8**, 167–176 (2017).

70. K. J. Friston, P. Rotshtein, J. J. Geng, P. Sterzer, R. N. Henson, A critique of functional localisers. *Neuroimage* **30**, 1077–1087 (2006).

71. K. J. Friston, P. Rotshtein, J. J. Geng, P. Sterzer, R. N. Henson, A Critique of Functional Localizers (2010) https:/doi.org/10.7551/mitpress/9780262014021.003.0002 (September 1, 2022).

72. S. J. Hanson, The Failure of Blobology: fMRI Misinterpretation, Maleficience and Muddle. *Front. Hum. Neurosci.* **16**, 870091 (2022).

73. S. Malik-Moraleda, *et al.*, An investigation across 45 languages and 12 language families reveals a universal language network. *Nat. Neurosci.* **25**, 1014–1019 (2022).

74. A. D. Friederici, S. M. E. Gierhan, The language network. *Curr. Opin. Neurobiol.* **23**, 250–254 (2013).

75. E. Fedorenko, S. L. Thompson-Schill, Reworking the language network. *Trends Cogn. Sci.* **18**, 120–126 (2014).

76. I. Voitov, T. D. Mrsic-Flogel, Cortical feedback loops bind distributed representations of working memory. *Nature* **608**, 381–389 (2022).

77. A. L. Barth, J. F. A. Poulet, Experimental evidence for sparse firing in the neocortex. *Trends Neurosci.* **35**, 345–355 (2012).

78. T. J. Ryan, C. Ortega-de San Luis, M. Pezzoli, S. Sen, Engram cell connectivity: an evolving substrate for information storage. *Curr. Opin. Neurobiol.* **67**, 215–225 (2021).

79. E. Fedorenko, C. Shain, Similarity of computations across domains does not imply shared implementation: The case of language comprehension. *Curr. Dir. Psychol. Sci.* **30**, 526–534 (2021).

80. R. Asano, C. Boeckx, K. Fujita, Moving beyond domain-specific versus domain-general options in cognitive neuroscience. *Cortex* **154**, 259–268 (2022).

81. L. Wehbe, *et al.*, Incremental Language Comprehension Difficulty Predicts Activity in the Language Network but Not the Multiple Demand Network. *Cereb. Cortex* **31**, 4006–4023 (2021).

82. I. Blank, N. Kanwisher, E. Fedorenko, A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *J. Neurophysiol.* **112**, 1105–1118 (2014).

83. C. Shain, I. A. Blank, M. van Schijndel, W. Schuler, E. Fedorenko, fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia* **138**, 107307 (2020).

84. Z. Mineroff, I. A. Blank, K. Mahowald, E. Fedorenko, A robust dissociation among the language, multiple demand, and default mode networks: Evidence from inter-region correlations in effect size. *Neuropsychologia* **119**, 501–511 (2018).

85. M. Assem, M. F. Glasser, D. C. Van Essen, J. Duncan, A Domain-General Cognitive Core Defined in Multimodally Parcellated Human Cortex. *Cereb. Cortex* **30**, 4361–4380 (2020).

86. M. L. Anderson, J. Kinnison, L. Pessoa, Describing functional diversity of brain regions and brain networks. *Neuroimage* **73**, 50–58 (2013).

87. D. Yin, *et al.*, A graph representation of functional diversity of brain regions. *Brain Behav.* **9**, e01358 (2019).

88. J. I. Skipper, J. T. Devlin, D. R. Lametti, The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain Lang.* **164**, 77–105 (2017).

89. M. L. Anderson, Précis of After Phrenology: Neural Reuse and the Interactive Brain. *Behav. Brain Sci.*, 1–22 (2015).

90. S. Dehaene, L. Cohen, Cultural recycling of cortical maps. *Neuron* **56**, 384–398 (2007).

91. M. L. Anderson, Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* **33**, 245–66; discussion 266–313 (2010).

92. M. Healey, *et al.*, More Than Words: Extra-Sylvian Neuroanatomic Networks Support Indirect Speech Act Comprehension and Discourse in Behavioral Variant Frontotemporal Dementia. *Front. Hum. Neurosci.* **14**, 598131 (2020).

93. I. Hertrich, S. Dietrich, C. Blum, H. Ackermann, The Role of the Dorsolateral Prefrontal Cortex for Speech and Language Processing. *Front. Hum. Neurosci.* **15**, 645209 (2021).

94. X. Jiang, “Prefrontal Cortex: Role in Language Communication during Social Interaction” in *Prefrontal Cortex*, (unknown, 2018) (September 1, 2022).

95. Kurby, Zacks, Situation models in naturalistic comprehension. *neuroscience of natural language use* (2015).

96. Ferstl, Neuroimaging of text comprehension: Where are we now. *Riv. Linguist.* (2010).

97. U. Hasson, H. C. Nusbaum, S. L. Small, Task-dependent organization of brain regions active during rest. *Proceedings of the National Academy of Sciences* **106**, 10841–10846 (2009).

98. H. Song, B.-Y. Park, H. Park, W. M. Shim, Cognitive and Neural State Dynamics of Narrative Comprehension. *J. Neurosci.* **41**, 8972–8990 (2021).

99. E. M. Gordon, *et al.*, Default-mode network streams for coupling to language and control systems. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 17308–17319 (2020).

100. J. I. Skipper, A voice without a mouth no more: The neurobiology of language and consciousness. *Neurosci. Biobehav. Rev.* **140**, 104772 (2022).

101. V. Menon, 20 years of the default mode network: A review and synthesis. *Neuron* (2023)

https:/doi.org/10.1016/j.neuron.2023.04.023.

102. D. W. Green, J. Abutalebi, Language control in bilinguals: The adaptive control hypothesis. *J. Cogn. Psychol.* **25**, 515–530 (2013).

103. M. Belyk, S. Brown, Perception of affective and linguistic prosody: an ALE meta-analysis of neuroimaging studies. *Soc. Cogn. Affect. Neurosci.* **9**, 1395–1403 (2014).

104. S. Frühholz, W. Trost, D. Grandjean, The role of the medial temporal limbic system in processing emotions in voice and music. *Prog. Neurobiol.* **123**, 1–17 (2014).

105. A. Z. Durfee, S. M. Sheppard, M. L. Blake, A. E. Hillis, Lesion loci of impaired affective prosody: A systematic review of evidence from stroke. *Brain Cogn.* **152**, 105759 (2021).

106. I. Rangus, M. Fritsch, M. Endres, B. Udke, C. H. Nolte, Frequency and phenotype of thalamic aphasia. *J. Neurol.* (2021) https:/doi.org/10.1007/s00415-021-10640-4.

107. B. Crosson, The Role of Cortico-Thalamo-Cortical Circuits in Language: Recurrent Circuits Revisited. *Neuropsychol. Rev.* (2019) https:/doi.org/10.1007/s11065-019-09421-8.

108. C. Boeckx, A. Benítez-Burraco, The shape of the human language-ready brain. *Front. Psychol.* **5**, 282 (2014).

109. J. I. Skipper, D. R. Lametti, Speech Perception under the Tent: A Domain-general Predictive Role for the Cerebellum. *J. Cogn. Neurosci.* **33**, 1517–1534 (2021).

110. Y. N. Yoncheva, J. D. Zevin, U. Maurer, B. D. McCandliss, Auditory Selective Attention to Speech Modulates Activity in the Visual Word Form Area. *Cereb. Cortex* **20**, 622–632 (2009).

111. K. von Kriegstein, *et al.*, Simulation of talking faces in the human brain improves auditory speech recognition. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 6747–6752 (2008).

112. B. Mathias, *et al.*, Visual Sensory Cortices Causally Contribute to Auditory Word Recognition Following Sensorimotor-Enriched Vocabulary Training. *Cereb. Cortex* **31**, 513–528 (2021).

113. M. D. Hauser, N. Chomsky, W. T. Fitch, The faculty of language: What is it, who has it, and how did it evolve? *Science* **298**, 1569–1579 (2002).

114. T. Q. Gentner, K. M. Fenn, D. Margoliash, H. C. Nusbaum, Recursive syntactic pattern learning by songbirds. *Nature* **440**, 1204–1207 (2006).

115. F. Mollica, *et al.*, Composition is the Core Driver of the Language-selective Network. *Neurobiology of Language* **1**, 104–134 (2020).

116. J. I. Skipper, H. C. Nusbaum, S. L. Small, “Lending a helping hand to hearing: Another motor theory of speech perception” in *Action to Language via the Mirror Neuron System*, M. A. Arbib, Ed. (Cambridge University Press, 2006), pp. 250–286.

117. M. Rombach, M. Porter, J. Fowler, P. Mucha, Core-Periphery Structure in Networks. *SIAM J. Appl. Math.* **74**, 167–190 (2014).

118. P. Csermely, A. London, L.-Y. Wu, B. Uzzi, Structure and dynamics of core/periphery networks. *J Complex Netw* **1**, 93–123 (2013).

119. M. Cinelli, G. Ferraro, A. Iovanella, Resilience of Core-Periphery Networks in the Case of Rich-Club. *Complexity* **2017** (2017).

120. X. Shen, S. Aliko, Y. Han, J. I. Skipper, C. Peng, Finding Core-Periphery Structures With Node Influences. *IEEE Transactions on Network Science and Engineering* **9**, 875–887 (2022).

121. W. James, The principles of psychology, Vol I (1890) https:/doi.org/10.1037/10538-000.

122. N. A. Crossley, *et al.*, The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* **137**, 2382–2395 (2014).

123. J. S. Siegel, *et al.*, Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E4367–76 (2016).

124. J. C. Griffis, R. Nenert, J. B. Allendorfer, J. P. Szaflarski, Damage to white matter bottlenecks contributes to language impairments after left hemispheric stroke. *Neuroimage Clin* **14**, 552–565 (2017).

125. J. C. Griffis, N. V. Metcalf, M. Corbetta, G. L. Shulman, Structural Disconnections Explain Brain Network Dysfunction after Stroke. *Cell Rep.* **28**, 2527–2540.e9 (2019).

126. C. Gratton, E. M. Nomura, F. Pérez, M. D’Esposito, Focal brain lesions to critical locations cause widespread disruption of the modular organization of the brain. *J. Cogn. Neurosci.* **24**, 1275–1285 (2012).

127. D. E. Warren, *et al.*, Network measures predict neuropsychological outcome after brain injury. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14247–14252 (2014).

128. M. C. Brady, H. Kelly, J. Godwin, P. Enderby, P. Campbell, “Speech and language therapy for aphasia following stroke” in *Cochrane Database of Systematic Reviews*, (John Wiley & Sons, Ltd, 2016).

129. M. L. Seghier, C. J. Price, Interpreting and Utilising Intersubject Variability in Brain Function. *Trends Cogn. Sci.* **22**, 517–530 (2018).

130. E. S. Finn, *et al.*, Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *Neuroimage* **215**, 116828 (2020).

131. J. Dubois, R. Adolphs, Building a Science of Individual Differences from fMRI. *Trends Cogn. Sci.* **20**, 425–443 (2016).

132. K. J. Michon, D. Khammash, M. Simmonite, A. M. Hamlin, T. A. Polk, Person-specific and precision neuroimaging: Current methods and future directions. *Neuroimage* **263**, 119589 (2022).

133. T. Yarkoni, R. A. Poldrack, T. E. Nichols, D. C. Van Essen, T. D. Wager, NeuroSynth: a new platform for large-scale automated synthesis of human functional neuroimaging data (2011).

134. A. R. Laird, J. L. Lancaster, P. T. Fox, BrainMap: the social evolution of a human brain mapping database. *Neuroinformatics* **3**, 65–78 (2005).

135. S. B. Eickhoff, *et al.*, Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Hum. Brain Mapp.* **30**, 2907–2926 (2009).

136. P. T. Fox, J. L. Lancaster, Mapping context and content: the BrainMap model. *Nat. Rev. Neurosci.* **3**, 319–321 (2002).

137. P. T. Fox, *et al.*, BrainMap taxonomy of experimental design: description and evaluation. *Hum. Brain Mapp.* **25**, 185–198 (2005).

138. R. W. Cox, AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).

139. J. Dockès, *et al.*, NeuroQuery, comprehensive meta-analysis of human brain mapping. *Elife* **9** (2020).

140. B. Fischl, FreeSurfer. *Neuroimage* **62**, 774–781 (2012).

141. C. Destrieux, B. Fischl, A. Dale, E. Halgren, Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* **53**, 1–15 (2010).

142. S. Aliko, J. Huang, F. Gheorghiu, S. Meliss, J. I. Skipper, A naturalistic neuroimaging database for understanding the brain using ecological stimuli. *Sci Data* **7**, 347 (2020).

143. D. Lynott, L. Connell, M. Brysbaert, J. Brand, J. Carney, The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behav. Res. Methods* **52**, 1271–1291 (2020).

144. L. Snoek, S. Miletić, H. S. Scholte, How to control for confounds in decoding analyses of neuroimaging data. *Neuroimage* **184**, 741–760 (2019).

145. E. Fedorenko, J. Duncan, N. Kanwisher, Language-selective and domain-general regions lie side by side within Broca’s area. *Curr. Biol.* **22**, 2059–2062 (2012).

146. J. A. van Dijk, A. Fracasso, N. Petridou, S. O. Dumoulin, Linear systems analysis for laminar fMRI: Evaluating BOLD amplitude scaling for luminance contrast manipulations. *Sci. Rep.* **10**, 5462 (2020).

147. B. McFee, *et al.*, Librosa Audio and Music Signal Analysis in Python in (2015) (November 29, 2021).

148. Y. Uranishi, OpenCV: Open Source Computer Vision Library. *The Journal of The Institute of Image Information and Television Engineers* **72**, 736–739 (2018).

149. W. J. B. van Heuven, P. Mandera, E. Keuleers, M. Brysbaert, SUBTLEX-UK: a new and improved word frequency database for British English. *Q. J. Exp. Psychol.* **67**, 1176–1190 (2014).

150. G. Chen, Z. S. Saad, J. C. Britton, D. S. Pine, R. W. Cox, Linear mixed-effects modeling approach to FMRI group analysis. *Neuroimage* **73**, 176–190 (2013).

151. R. W. Cox, Equitable Thresholding and Clustering: A Novel Method for Functional Magnetic Resonance Imaging Clustering in AFNI. *Brain Connect.* **9**, 529–538 (2019).

152. R. W. Cox, AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).

153. P. Barucca, D. Tantari, F. Lillo, Centrality metrics and localization in core-periphery

networks. *arXiv [cs.SI]* (2015).

154. Q. K. Telesford, S. L. Simpson, J. H. Burdette, S. Hayasaka, P. J. Laurienti, The brain as a complex system: using network science as a tool for understanding the brain. *Brain Connect.* **1**, 295–308 (2011).

155. S. Oldham, *et al.*, Consistency and differences between centrality measures across distinct classes of networks. *PLoS One* **14**, e0220061 (2019).

156. C. Li, Q. Li, P. Van Mieghem, H. E. Stanley, H. Wang, Correlation between centrality metrics and their application to the opinion model. *Eur. Phys. J. B* **88**, 65 (2015).

157. C. F. Beckmann, S. M. Smith, Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* **23**, 137–152 (2004).