Check for updates

The MIT Press

RESEARCH ARTICLE

Brain Areas Critical for Picture Naming: A Systematic Review and Meta-Analysis of Lesion-Symptom Mapping Studies

Vitória Piai[1,2] (iD) and Dilys Eikelboom[3]

[1]Radboud University, Donders Centre for Cognition, Nijmegen, Netherlands
[2]Radboudumc, Donders Centre for Medical Neuroscience, Department of Medical Psychology, Nijmegen, Netherlands
[3]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

ABSTRACT

Lesion-symptom mapping (LSM) studies have revealed brain areas critical for naming, typically finding significant associations between damage to left temporal, inferior parietal, and inferior fontal regions and impoverished naming performance. However, specific subregions found in the available literature vary. Hence, the aim of this study was to perform a systematic review and meta-analysis of published lesion-based findings, obtained from studies with unique cohorts investigating brain areas critical for accuracy in naming in stroke patients at least 1 month post-onset. An anatomic likelihood estimation (ALE) meta-analysis of these LSM studies was performed. Ten papers entered the ALE meta-analysis, with similar lesion coverage over left temporal and left inferior frontal areas. This small number is a major limitation of the present study. Clusters were found in left anterior temporal lobe, posterior temporal lobe extending into inferior parietal areas, in line with the arcuate fasciculus, and in pre- and postcentral gyri and middle frontal gyrus. No clusters were found in left inferior frontal gyrus. These results were further substantiated by examining five naming studies that investigated performance beyond global accuracy, corroborating the ALE meta-analysis results. The present review and meta-analysis highlight the involvement of left temporal and inferior parietal cortices in naming, and of mid to posterior portions of the temporal lobe in particular in conceptual-lexical retrieval for speaking.

INTRODUCTION

According to psycholinguistic models of language production, a speaker starts with a concept they want to express and goes through several stages until their intention can be articulated. Generally speaking, these stages can be seen as conceptual preparation, lexical selection (i.e., an operation at the level of "lemmas," a semantic-syntactic representation), phonological retrieval and encoding (i.e., the retrieval and ordering of the speech sounds associated with that lemma), phonetic encoding (i.e., the computation of the gestural score), and articulation (e.g., Dell, 1986; Dell & O'Seaghdha, 1992; Levelt et al., 1999).

Producing language involves an extensive network of brain areas. Neurolinguistic models of language production have attempted to link the proposed cognitive stages to different brain areas (Hickok & Poeppel, 2007; Indefrey & Levelt, 2004; Roelofs, 2014). Various methods have been used to uncover these neural substrates, for example, by combining word

production tasks with functional neuroimaging (e.g., Price et al., 1996), electrophysiology (e.g., Liljeström et al., 2009), or neurostimulation techniques (e.g., Hernandez-Pavon et al., 2014). In general, these methods have highlighted the correlation between activity in a broad fronto-temporo-parietal network and word production.

A different, noncorrelational approach to studying language production consists of examining the consequences of tissue damage on performance, as done with lesion-symptom mapping (LSM) techniques. Although differing in the particular methodology, these techniques aim to map the relationship between lesions and behavior, generating statistical maps. Rather than requiring a cut-off score or binary data (e.g., behavioral performance is either deficient or not), a continuous behavioral score can be used, thus enabling a more sensitive approach. Furthermore, these techniques can serve as a whole-brain analysis, rather than studying particular regions-of-interest (ROIs), although they are inherently limited by lesions coverage. In one particular and much used approach, voxel-based lesion-symptom mapping (VLSM), a statistical test is run at every voxel, comparing a behavioral score between patients with and without a lesion in that voxel, thus identifying voxels critical for the measured behavioral performance (Bates et al., 2003). Besides VLSM, similar (and more recent) lesion-symptom mapping approaches exist, such as voxel-based morphometry (Ashburner & Friston, 2000), voxel-based correlational methodology (VBCM; Tyler et al., 2005), and multivariate methods (reviewed in Ivanova et al., 2021), such as support vector regression multivariate LSM (Zhang et al., 2014). We note that a discussion of how these methods work is beyond the scope of the present study (see for explanation and comparisons between these methods, e.g., Geva et al., 2012; Ivanova et al., 2021).

Baldo et al. (2013) applied VLSM on naming accuracy in chronic stroke patients, while controlling for overall fluency in speech production and visual recognition of the items in the naming test. Significant brain regions were predominantly found in the left mid and posterior portions of the middle temporal gyrus (MTG), suggesting that naming critically depends on this area and the adjacent white matter. A different study by Thye and Mirman (2018) also found that damage to the left MTG was associated with deficits in naming in stroke patients, in addition to areas in the left inferior frontal gyrus (IFG), supramarginal gyrus (SMG), and angular gyrus (AG). Similar LSM studies confirmed the involvement of these areas but also found other or additional areas associated with naming performance, such as the left postcentral gyrus, inferior temporal gyrus (ITG), inferior longitudinal fasciculus, and temporal pole (e.g., Alyahya et al., 2018b; Faroqi-Shah et al., 2014; Piras & Marangolo, 2007).

Present Study

In sum, damage to left temporal, inferior parietal (AG and SMG) and inferior frontal areas is typically associated with deficits in naming, though specific subregions found in the literature vary. An informal attempt to summarize available LSM evidence may be complicated by the fact that comparability is limited when different studies implement different experimental designs. Studies may for example vary in employed LSM approach, task demand, and covariates used. Furthermore, small sample sizes may be investigated, resulting in lower reliability. However, the use of a formal meta-analytic approach allows for a quantitative review of a large body of LSM data, enabling the identification of locations in the brain that show consistent relationships to behavior across studies (Eickhoff et al., 2012). Hence, the aim of the present study was to perform a systematic review and an anatomic likelihood estimation (ALE) meta-analysis of studies using LSM methods in combination with a naming task, to identify a pattern of consistent associations between brain lesions and word production. In addition, a more in-depth analysis was performed attempting to align the processing stages most likely tapped into by a meta-analyzed

study to stages proposed by psycholinguistic models of word production (e.g., Dell, 1986; Dell & O'Seaghdha, 1992; Indefrey & Levelt, 2004; Levelt et al., 1999).

In our systematic review, papers using any form of LSM were considered. Papers had to implement oral naming as a behavioral task in some form. To limit heterogeneity across the researched participant group, we only included studies in individuals beyond the acute stages of stroke (here defined as at least one month post-stroke). Papers in which the dependent variable was global accuracy in naming performance were considered for global accuracy analysis. From these, only papers that provided coordinates qualified for ALE meta-analysis (see Figure 1). Papers with a dependent variable more elaborate than global accuracy, such as error type in naming or combining naming with another language task, were not considered for the ALE meta-analysis since there was not enough consistency across them, which would introduce large heterogeneity in the dependent variable tested. Instead, these papers were considered for a beyond-accuracy analysis in narrative form. Dependent variables from these papers were linked to the stages of word production described above in an attempt to elucidate the ALE meta-analysis results.

MATERIALS AND METHODS

Literature Search and Selection

A systematic search was performed, using the Web of Science Core Collection (Clarivate, 2023) and APA PsycINFO (American Psychological Association, 2023) databases using the
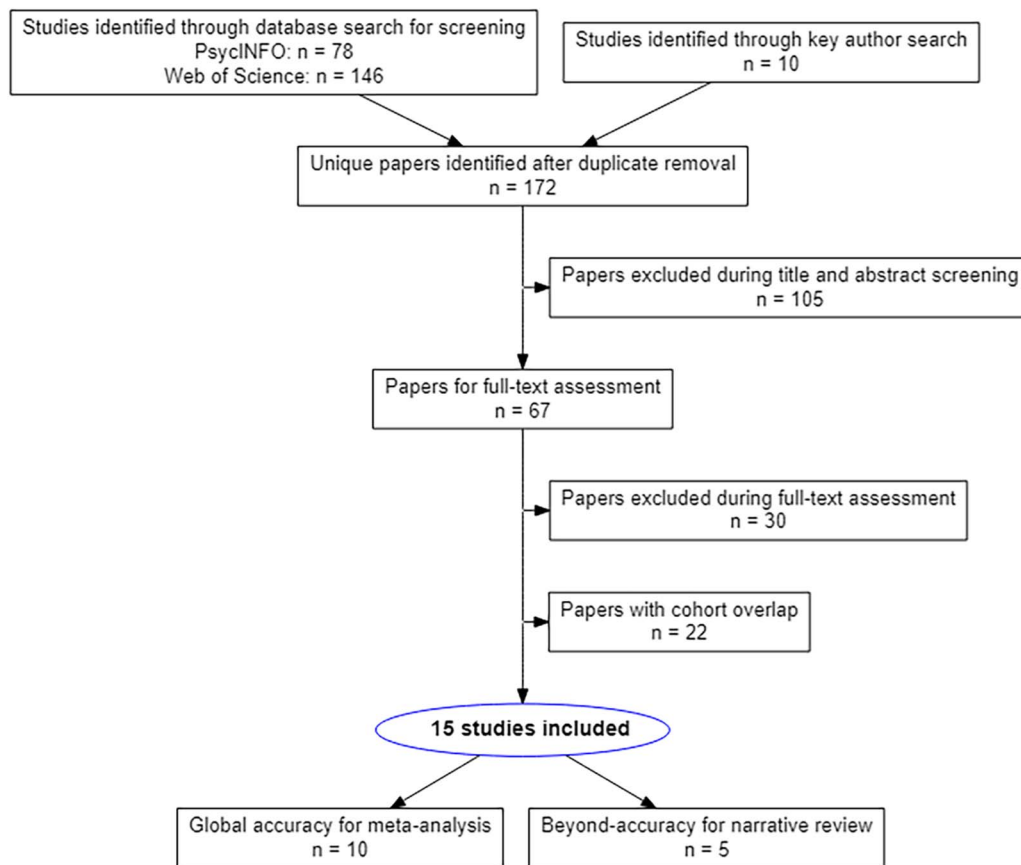
Figure 1. PRISMA flow diagram of the literature selection process.

following keywords: [("naming" OR "production") AND ("stroke" OR "poststroke" OR "infarct" OR "CVA" OR "cerebral vascular accident*" OR "cerebrovascular accident*" OR "post stroke") AND ("lesion behavio*r mapping" OR "voxel wise" OR "voxel based" OR "symptom mapping" OR "lesion mapping")]. This resulted in a list of 162 unique papers (last updated on September 18, 2021). To identify other papers not picked up in the automatic search, publications from the following authors (who have published seminal studies with LSM methodology) were screened: Baldo, J.; Binder, J. R.; Crinion, J. T.; Dragoy, O.; Fridriksson, J.; Hope, T. M. H.; Lambon Ralph, M. A.; Mirman, D.; Price, C. J.; Schwartz, M. E.; and Wilson, S. M. Furthermore, the reference list of a review paper by Mirman and Thye (2018) was searched for relevant papers. This yielded 10 additional papers. A PRISMA flowchart showing the selection of these papers can be found in Figure 1.

Papers were independently screened by two authors (title and abstract in phase one, full text in phase two) on whether they satisfied the following selection criteria: papers had to (1) be original empirical work; (2) include stroke patients only; (3) state clearly that all patients were at least one month post onset of stroke; (4) involve LSM of (5) oral picture naming task performance; (6) not reporting single cases; (7) not based on functional imaging or using synthetic data; (8) not studying predefined ROIs. During full-text screening, papers (with corresponding total number) were excluded according to the following criteria: Not single-word oral picture naming ($N = 13$); no LSM methodology ($N = 6$); not all patients >1 month post-stroke ($N = 4$); not (only) stroke cohort ($N = 2$); no coordinates provided ($N = 3$); no original or real data ($N = 2$).

Paper Categorization

Papers were categorized according to the dependent variable used for LSM, resulting in a global accuracy analysis data set (i.e., accuracy in naming performance, not further specified). An exception was made for papers studying a compound naming score from the Western Aphasia Battery (Kertesz, 1982). This score consists of four naming subtests, one of which is oral naming, which we considered to suffice for our analyses, thus allowing us to include more papers in the ALE meta-analysis. Papers in the global accuracy analysis data set had to provide coordinates to be included for ALE meta-analysis. Authors were contacted for additional information on foci coordinates. Papers analyzing a more specific score, for example, the number of semantic or phonological errors or after a dimensionality reduction step (e.g., using principal component analysis [PCA]), were included for a beyond-accuracy analysis.

Next, for each type of analysis (i.e., global accuracy or beyond-accuracy), relevant papers were screened for potential overlap in the participant sample by checking overlap in authors and noting where participants were recruited from. For unclear cases, authors were contacted to gain information regarding the participants tested, but no response was received. To reduce the risk of duplicated data, which inflates the effect size, from every subgroup of potentially overlapping papers, the paper that suited our research purpose best was selected in the following way. If dependent variables used in the overlapping papers were equal, the paper with the largest cohort was selected. Priority was given to papers providing coordinates over papers without them. Selection priority was also given to the dependent variable reflecting the most clear-cut measure of naming (as opposed to the naming score being used with a technique for dimension reduction such as PCA, which is then related to lesion information), noun naming specifically rather than verb naming (since verb naming was much less common across the studies). If multiple LSM analyses were performed within one paper, selection priority was given to results from analyses of which noun naming was the largest part, as this is the most commonly reported measure of naming across papers. Also, univariate VLSM results were

prioritized over any other form of LSM, and so were manually rather than automatically traced lesion maps, as these two types tend to be the most common approaches. Furthermore, priority was given to results controlling for all available covariates. The choice for the most commonly used approach was meant to increase the comparability across studies and, thus, decrease heterogeneity.

For the ALE meta-analysis, clusters of (potentially) overlapping papers were found. Out of the first cluster (Alyahya et al., 2018a, 2018b; Butler et al., 2014), the paper by Alyahya et al. (2018b) was selected for including noun naming (cf. Alyahya et al., 2018a) and a more suitable dependent variable to our study (cf. Butler et al., 2014). The second cluster consisted of two papers (Piras & Marangolo, 2007, 2010), for which Piras and Marangolo (2010) provided coordinates. For the third cluster (Pustina et al., 2016; Thye & Mirman, 2018; Zhang et al., 2014), the paper by Thye and Mirman (2018) was selected given the availability of foci coordinates. For the two papers by Lukic et al. (2017, 2021), using verb naming, the one with the larger sample size was chosen (Lukic et al., 2021).

For beyond-accuracy analysis, clusters of (potentially) overlapping papers were obtained. The first cluster contained three papers (Fridriksson et al., 2016, 2018; Stark et al., 2019), out of which Fridriksson et al. (2018) was selected for studying different error types within naming (cf. Fridriksson et al., 2016) and the larger sample size (cf. Stark et al., 2019). The second cluster consisted of nine papers (Alyahya et al., 2018a, 2020a, 2020b; Butler et al., 2014; Halai et al., 2017, 2018a, 2018b; Tochadse et al., 2018; Zhao et al., 2020), out of which Tochadse et al. (2018) was selected. In this paper, semantic and phonological errors were studied, yielding s and p parameters, respectively, according to the computational model of Dell (Dell, 1986; Dell et al., 2013), whereas all other papers in this cluster used PCA on a neuropsychological test battery, except for Halai et al. (2018b). This latter paper studied different error types within naming but was excluded as Tochadse et al. (2018) provided, in our opinion, a theoretically better motivated distinction between error types than Halai et al. (2018b). The final cluster consisted of nine papers (Chen et al., 2019; Dell et al., 2013; Mirman, Chen, et al., 2015; Mirman & Graziano, 2013; Mirman, Zhang, et al., 2015; Schwartz et al., 2009, 2011, 2012; Walker et al., 2011), of which Dell et al. (2013) was chosen for conducting VLSM on parameters derived from computational modeling (as in Tochadse et al., 2018).

The selection yielded 15 original research papers that used LSM techniques in combination with a (picture) naming task in stroke individuals >1 month post-onset with non-overlapping cohorts. Ten papers qualified for ALE meta-analysis of global accuracy and five for beyond-accuracy analysis. The selection and categorization procedure of the included papers can be found in a PRISMA flowchart, shown in Figure 1.

Quality Assessment

To try to chart the heterogeneity across studies included in the meta-analysis, we performed a quality assessment of the evidence for the purpose of our systematic review by checking various parameters. We note that this does not speak to the quality of the papers themselves, but rather to the quality of the evidence as it impacts our synthesis and findings. Papers could receive a maximum of 6 points on the (clarity of the description of the) studied population, 2 points for the clarity of the description of the task and how performance was scored, 4 points for the (description of the) statistical analysis, and finally 1 point for the clarity of the outcome measure, with a maximum of 13 points in total. Details regarding the parameters and weighted distributions of points can be found in Table S1 in the Supporting Information, available at https://doi.org/10.1162/nol_a_00097, and scoring per paper can be found in Table S2.

ALE Meta-analysis

ALE meta-analysis was performed using the revised version implemented in BrainMap Ginger-ALE 3.0 software following the MNI152 template (Eickhoff et al., 2009, 2012; Turkeltaub et al., 2012), in conjunction with anatomical data (rather than using functional neuroimaging data for activation likelihood estimation meta-analyses, see for a similar approach Na et al., 2022; Urgesi et al., 2014). These data concerned peak coordinates of significant clusters associated with naming task performance as a result of LSM. All extracted coordinates were reported in Montreal Neurological Institute (MNI) space.

In ALE analysis, foci from a given study or experiment are modeled as Gaussian probability density distributions with a full-width half-maximum (FWHM) calculated from the experiment's sample size and merged together to form a map. This map therefore represents a summary of the results of that study, taking into account between-subject and between-template variability (e.g., caused by data smoothing and standardization into anatomical space), by modeling foci as probability distributions rather than singular points. These probability distributions are then taken together by calculating the voxel-wise union of the maps from different studies, to assign to every voxel an ALE value equal to the probability that at least one of the foci in the data set actually lies within this voxel (Turkeltaub et al., 2002, 2012). Lastly, convergence of foci across experiments is tested by comparing the calculated ALE values against ALE values obtained under an empirically defined null distribution reflecting random spatial association. A whole-brain map can then be produced, showing the differential likelihood of associations (in our case, between lesion and naming score) at all brain locations afforded by the lesion coverage. Significance was assessed using a cluster-level familywise error correction set at $p < 0.05$, with a cluster forming threshold set at $p < 0.01$ and 1000 permutations. Anatomical labels were obtained from BrainMap GingerALE 3.0, based on the Talairach Daemon (1988 Talairach atlas). We note that the use of a Gaussian probability density distribution with FWHM, which is commonly used for functional magnetic resonance imaging (fMRI) studies, may not be the best option for an ALE meta-analysis of LSM studies. However, fMRI is a haemodynamic measure shaped by properties of the vascular system and strokes are vascular in nature, motivating the use of this distribution. This issue remains nevertheless a limitation of our approach, as no empirical studies exist validating the use of this probability distribution in ALE meta-analyses for LSM data.

RESULTS

ALE Meta-analysis of Global Accuracy

Descriptions of the 10 papers used for ALE meta-analysis studying accuracy in naming are shown in Table 1. These papers in total regarded 69 foci, acquired from 534 subjects. As far as we could establish, within each paper coverage over left temporal and left inferior frontal areas was similar; as such, there was in general no particular bias to frontal cortex relative to temporal cortex.

Quality of the evidence was assessed and total score per paper can be found in Table 1. Papers scored 9 or higher. Detailed scoring per quality parameter can be found in Table S2.

Results of the ALE analysis can be found in Figure 2 and Table 2. Here, we follow the subdivision of the temporal lobe into anterior, mid, and posterior portions by Indefrey and Levelt (2004), with corresponding boundary y coordinates in Talairach space at −7 and −38. Four significant clusters were identified. Cluster 1, with 7 peaks contributed by seven studies, had the maximal ALE value (ALE = 0.028) in left anterior temporal cortex (MNI −42, −2,

Table 1. List of papers included for anatomic likelihood analysis meta-analysis, studying global accuracy in naming.

Study	N (#f)[c]	Age (range or M ± SD)	Post-stroke time (months)	Language	Dependent variable	Lesion delineation	Modality	Analysis	Lesion volume as covariate	Min. subjects with lesioned voxel[d]	No. of foci (and cluster contribution)[e]	Quality (/13)
Akinina et al., 2019	40 (21)	33–78	>3	Russian	Verb naming accuracy (picture)	Manual	MRI	VLSM	Yes	10% (n = 4)	1 (1)	12
Alyahya et al., 2018b	48 (14)	44–87	>12	English	Noun naming accuracy (picture)	Automated	MRI	VBCM	Yes	n.m.	8 (1, 2)	12
Baldo et al., 2013	96 (21)	31–84	>3	English	Noun naming accuracy (picture)	Manual	MRI, CT	VLSM	n.m.	5% (n = 5)	1 (2)	12
Faroqi-Shah et al., 2014	31 (10)	42–72	>10	English	Noun naming accuracy (picture)	Manual	MRI	VLSM	n.m.	13% (n = 4)	5 (2, 3)	11
Geva et al., 2012[a]	21 (7)	21–81	>6	English	Noun naming accuracy (CAT)	Manual	MRI	VLSM	Yes	5% (n = 1)	12* (2, 3, 4)	12
Griffis et al., 2017	43 (18)	23–90	>12	English	Noun naming accuracy (picture)	Automated	MRI	SVR-LSM	Yes	23% (n = 10)	13 (1, 2, 3)	12
Lukic et al., 2021	76 (26)	22–81	>8	English	Verb naming accuracy (NNB)	Semiautomated	MRI	VLSM	Yes	10% (n = 7)	2 (1, 3)	13
Piras & Marangolo, 2010	20 (7)	38–78	>6	Italian	Noun naming accuracy (picture)	Manual	MRI	VLSM	Yes	25% (n = 5)	5 (1, 3)	13

Sul et al., 2019	31 (15)	55.5 ± 11.5	>12	Korean	Naming score (K-WAB)	Manual	MRI	VLSM	n.m.	10% (n = 3)	7 (1, 2)	9
Thye & Mirman, 2018[b]	128 (57)	26–79	>1	English	Noun naming accuracy (PNT)	Manual	MRI, CT	VLSM	Yes	10% (n = 13)	15* (1, 3, 4)	10

Note. Post-stroke time was the minimum time between stroke onset and scanning or testing, whichever was performed earlier. Quality assessment was performed by scoring different parameters out of a maximum of 13 points (details and weighted distributions of points can be found in Table S1, scoring per paper can be found in Table S2). CAT = Comprehensive Aphasia Test; NNB = Northwestern Naming Battery; K-WAB = Korean version of the Western Aphasia Battery; PNT = Philadelphia Naming Test; MRI = magnetic resonance imaging; CT = computed tomography; VLSM = voxel-based lesion-symptom mapping; VBCM = voxel-based correlational methodology; SVR-LSM = support vector regression multivariate lesion-symptom mapping; n.m. = not mentioned in paper or supplementary material.

[a] For Geva et al. (2012), the statistical map was cluster thresholded at z > 4.61 by the first author of that study.

[b] For Thye and Mirman (2018), from the statistical map made available, center coordinates of clusters with >10 voxels were selected by the authors, in a procedure blinded for cluster/voxel location.

[c] Number of subjects with the amount of females stated in brackets.

[d] Minimum number of subjects with lesion in a specific voxel before this voxel is included in statistical analysis, presented as percentage out of the full cohort.

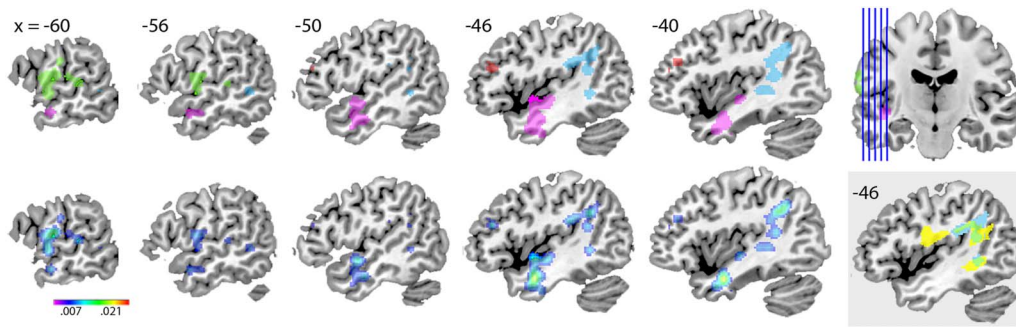[e] Number of significant foci obtained.

Figure 2. Results of the anatomic likelihood estimate (ALE) analysis. (Top) Location of the four clusters (cluster 1 in pink, cluster 2 in light blue, cluster 3 in green, cluster 4 in red), with significant association between global accuracy in naming task performance and brain lesions. Location of all sagittal slices is indicated in the upper right corner. (Bottom) ALE maps of the four clusters, corrected for cluster-level familywise error at an alpha level of 0.05 (following voxel-level threshold of 0.01). The color bar indicates the ALE value range. (Gray inset) Cluster 2 (in blue), arcuate fasciculus (in yellow), and their overlap (in green). The arcuate fasciculus mask was obtained from the Natbrain atlas (Catani & Thiebaut de Schotten, 2008).

−24), closest to MTG. Cluster 2, with 10 peaks contributed by six studies, had the maximal ALE value (ALE = 0.016) in left inferior parietal lobule (labeled angular gyrus in the AAL atlas; Devenyi et al., 2017). Cluster 3, with 5 peaks contributed by six studies, had the maximal ALE value (ALE = 0.017) in left postcentral gyrus. Cluster 4, with 5 peaks contributed by two studies, had the maximal ALE value (ALE = 0.011) in left.

Beyond-Accuracy Analysis

Details of the five papers considered for beyond-accuracy analysis can be found in Table 3. These papers measured different dependent variables. We tried to link their word production measure to proposed stages of word production (Dell, 1986; Levelt et al., 1999; Schwartz et al., 2004). We note that on the basis of the outcome measures reported, conceptual preparation and lexical selection could not be distinguished.

Out of the papers included for beyond-accuracy analysis, five papers studied a measure most likely associated with the conceptual-lexical selection stage, and three papers also studied a measure most likely associated with the phonological encoding stage.

Fridriksson et al. (2018) studied semantic and phonological errors, linked to the conceptual-lexical and phonological code stages, respectively. Lesion-symptom mapping results for semantic errors revealed significant areas overlapping with cluster 2, the strongest predictor of semantic error production being lesions in the left "posterior" (authors' own terminology) MTG. No significant regions were found to be predictive of phonological errors made in naming (cf. Schwartz et al., 2012). Of note, our cluster 3 shows overlap with their LSM results for articulation rate.

A different way to examine lexical-semantic versus phonological stages is through formalization of a computational model. Based on Dell's computational model of word production (Dell, 1986; Dell et al., 2013), two relevant parameters are defined: (1) The s parameter, representing the connection weights between conceptual and lexical units ("lemma access"), and (2) the p parameter, the connection weights between the lexical and phonological units. From performance in picture naming and nonword repetition tests, Dell et al. (2013) derived s and p parameters, which were analyzed with VLSM. The s parameter was associated with left "anterior" (authors' own terminology) STG and MTG, left temporal pole, and left middle and inferior frontal gyri, overlapping with our four clusters. In addition, the association was also present

Table 2. Significant anatomic likelihood analysis clusters and corresponding MNI coordinates of the local maxima.

Cluster	Volume (mm³)	Label	ALE score	Z score	Coordinates		
					x	y	z
1	6,736	Left Superior Temporal Gyrus, BA 38	0.028	6.3	−42	−2	−24
		Left Superior Temporal Gyrus, BA 22	0.013	3.9	−48	−2	−8
		Left Sub-lobar Insula, BA 13	0.012	3.8	−44	−14	−6
		Left Inferior Temporal Gyrus, BA 20	0.010	3.5	−50	−10	−22
		Left Inferior Temporal Gyrus, BA 20	0.010	3.5	−44	−8	−34
		Left Inferior Temporal Gyrus, BA 21	0.010	3.4	−60	−8	−16
		Left Middle Temporal Gyrus, BA 21	0.008	2.9	−54	2	−16
2	6,208	Left Superior Temporal Gyrus, BA 39	0.016	4.5	−42	−48	32
		Arcuate Fasciculus*	0.011	3.6	−40	−46	14
		Arcuate Fasciculus*	0.011	3.6	−46	−48	0
		Left Sub-lobar Insula, BA 13	0.011	3.6	−46	−30	24
		Left Caudate Tail	0.010	3.6	−38	−40	2
		Left Sub-lobar Insula, BA 13	0.010	3.5	−44	−36	28
		Left Sub-lobar Insula, BA 13	0.010	3.3	−38	−40	26
		Left Transverse Temporal Gyrus, BA 41	0.009	3.2	−36	−38	12
		Left Caudate Tail	0.009	3.2	−38	−32	2
		Left Middle Temporal Gyrus, BA 22	0.008	2.9	−56	−48	2
3	3,816	Left Transverse Temporal Gyrus, BA 42	0.017	4.5	−62	−10	14
		Left Superior Temporal Gyrus, BA 22	0.012	3.8	−62	−6	0
		Left Postcentral Gyrus, BA 2	0.010	3.4	−62	−16	26
		Left Superior Temporal Gyrus, BA 42	0.010	3.4	−60	−30	8
		Left Postcentral Gyrus, BA 40	0.008	2.9	−62	−24	14
4	2,360	Left Middle Frontal Gyrus, BA 9	0.011	3.7	−30	40	20
		Left Middle Frontal Gyrus, BA 9	0.010	3.5	−34	36	26
		Left Middle Frontal Gyrus, BA 10	0.010	3.5	−36	44	18
		Left Middle Frontal Gyrus, BA 46	0.010	3.5	−46	34	20
		Left Middle Frontal Gyrus, BA 9	0.010	3.5	−42	36	26

Note. BA = Brodmann Area. Anatomical labeling provided by BrainMap GingerALE 3.0, based on the Talairach Daemon 1988 atlas. MNI = Montreal Neurological Institute.

* No gray matter found, anatomical label derived from the Natbrain atlas (Catani & Thiebaut de Schotten, 2008).

Table 3.   List of papers studying a beyond-accuracy measure of naming.

Study	*N* (#f)[a]	Post-stroke time (months)	Language	Dependent variable	Stage of word production
Dell et al., 2013	103 (44)	>1	English	Semantic and phonological parameter weights via computational modelling	Conceptual/lexical selection and phonological encoding respectively
Fridriksson et al., 2018	105 (n.m.)	>6	English	Semantic and phonological noun naming errors	Conceptual/lexical selection and phonological encoding respectively
Harvey & Schnur, 2015	15 (4)	>6	English	Semantic interference in noun naming	Conceptual/lexical selection
Schnur et al., 2009	12 (n.m.)	>10	English	Growth of semantic interference in noun naming	Conceptual/lexical selection
Tochadse et al., 2018	53 (n.m.)	>12	English	Semantic and phonological parameter weights via computational modelling	Conceptual/lexical selection and phonological encoding respectively

Note. Post-stroke time was the minimum time between stroke onset and scanning or testing, whichever was performed earlier. n.m. = not mentioned in paper or supporting information.

[a] Number of subjects with the number of females stated in brackets.

posteriorly, at the temporo-parietal and parietal-temporal-occipital junctions, including AG. The *p* parameter was mainly associated with left SMG and postcentral gyrus (also including precentral gyrus and insula, similar to our cluster 3). Tochadse et al. (2018) similarly studied naming through *s* and *p* weights from Dell's computational model. Based on the peak coordinates obtained in their VBCM analysis and coordinates for the subdivision of the temporal lobe (based on Indefrey & Levelt, 2004), the *s* parameter was associated with regions in the mid portion of the left temporal lobe. Peak coordinates associated with the *p* parameter were located either in the anterior or mid portions of the temporal lobe.

Two other papers combined LSM with semantic interference, which could be linked to the conceptual-lexical selection stages of word production (Roelofs, 2018). Harvey and Schnur (2015) studied the areas involved in both semantic interference and growth of interference across cycles in naming. The largest significant cluster associated with semantic interference was located in the left posterior MTG (according to the subdivision adopted here: MNI −52, −40, −5, Talairach *y* = −40), close to our cluster 2, whilst the other cluster was located in the left mid MTG (according to the subdivision adopted here: MNI −49, −21, −8, Talairach *y* = −22). No region was significantly associated with growth in interference across naming cycles. This latter dependent variable, that is, the growth of interference, was specifically studied in a study by Schnur et al. (2009). VLSM analysis revealed that growth of interference was significantly related to voxels only in the "posterior" (author's own terminology) left IFG.

To conclude the beyond-accuracy analysis, a tendency seems to be present across studies for deficits in the conceptual preparation and/or lexical retrieval stages to be associated with lesions in somewhat more mid to posterior temporal regions. Regarding the phonological code retrieval stage, since only two studies obtained statistically significant results (Dell et al., 2013; Tochadse et al., 2018) that were not converging, the evidence remains inconclusive.

DISCUSSION

To investigate the brain areas critical for word production, the present study quantitatively compared results from papers combining lesion-symptom mapping with global accuracy scores in naming, by performing an ALE meta-analysis. We identified four separate clusters. One cluster was predominantly in the anterior portion of the left temporal lobe, in STG, MTG, and ITG. The second cluster was predominantly in the posterior portion of the left temporal lobe including the inferior parietal lobule, mostly in white matter. An overlay of this cluster with the outline of the arcuate fasciculus indicated a large degree of overlap. The third cluster had a peak in postcentral gyrus and the fourth cluster in middle frontal gyrus. No peaks were identified in the left IFG. This distribution was found despite a similar lesion coverage over left inferior frontal as over left temporal lobe areas. In general, the quality of the evidence across studies was good for the purpose of our review and meta-analysis. The vast majority of the studies were conducted in English-speaking countries, with only three other languages represented in our sample.

Three papers could not be included in the global accuracy due to missing coordinates. The three strongest predictors of correct naming obtained by Fridriksson et al. (2018) were the left "posterior" (authors' own terminology) STG, AG, and SMG. Anterior portions of the temporal lobe were also found in this study, though these predictors were less strong. Pillay et al. (2017) provided a VLSM map of picture naming in their supplementary materials, which revealed significant areas (through visual inspection) in posterior portions of left STG, inferior parietal lobule, lateral frontal cortex, and insula. Finally, in the study by Skipper-Kallal et al. (2017), impairment in picture naming was associated with left posterior STG (authors' own terminology) in particular, but also with left AG, intraparietal sulcus, and parts of the pars triangularis in the IFG.

Since overall naming scores reflect a mixture of errors, it is difficult to relate the patterns found to particular stages of word production. For example, while Akinina et al. (2019) took care to try to isolate the lexical stage, and Baldo et al. (2013) reported their findings while covarying for visual perception and overall speech fluency deficits, the global accuracy measure in other studies is less specific to one or a couple of stages. Therefore, we also synthesized studies examining measures beyond global accuracy in an attempt to elucidate the patterns found by relating them as much as possible to particular stages of word production as stipulated by psycholinguistic models. We found tentative evidence that conceptual preparation and/or lexical selection are associated with lesions in somewhat more mid to posterior temporal lobe regions, whereas the evidence for phonological encoding was less consistent across studies.

In the course of publishing this work, another meta-analysis of lesion-symptom mapping studies was published focusing on various language tasks (Na et al., 2022). For naming, the authors found a cluster in the left parahippocampal gyrus and left mid STG (MNI −59, −11, 7, Brodman Area 22, Talairach $y = -12$). This cluster is in the proximity of cluster 3 we identified. However, unlike in our meta-analysis, the authors did not differentiate between phase of the stroke (acute, subacute, and chronic were all included) or performance measure (global accuracy as well as specific error types were included) in the analysis.

While two previous (semi-)systematic reviews and meta-analyses have provided evidence on the neural substrates of more specific stages of word production based on correlational measures (Indefrey & Levelt, 2004; Price, 2012), here we explicitly sought to provide causal evidence. The meta-analysis of Indefrey and Levelt (2004) has suggested that lexical selection is associated with left MTG (and the mid portion in particular), whereas phonological code

retrieval (part of phonological encoding) is associated with left posterior MTG and STG. Syllabification (the ordering of phonemes into syllables, part of phonological encoding) is associated with left posterior IFG and phonetic encoding and articulation mainly with bilateral ventral motor and sensory regions. Our results for global naming accuracy are partly in line with this proposal. The beyond-accuracy analysis, by contrast, provides tentative evidence in agreement with this proposal regarding the conceptual/lexical stages, which in our findings were related to more mid to posterior temporal lobe regions.

Our results concern stroke populations only. However, given that different pathologies have their intrinsic spatial biases—in the case of stroke, due to cerebrovascular organization—trying to bridge across pathologies would be fruitful for our understanding of the neural basis of language production. These studies are however scarcer, so no meta-analysis is possible given the current literature. Of note, converging evidence for the present results can be found. For example, in an LSM study examining presurgical brain tumor cases (Faulkner & Wilshire, 2020), conceptual/lexical selection (operationalized as semantic errors and omissions in picture naming and category fluency performance covaried for letter fluency performance) was associated with left posterior MTG and parts of the lateral occipital cortex and ITG. Phonological encoding (operationalized as accuracy on word and nonword repetition and effect of word length in picture naming) was associated with left posterior SMG and AG. In individuals with Alzheimer disease, correlations between hypometabolism and naming errors have also been found (Isella et al., 2020). In particular, semantic errors were related to the mid portion of the MTG and left ITG. Formal errors (i.e., the resulting existing word resembles the target word in terms of its form, "rat" instead of "mat"), tapping both into the connections between lexical and phonological units and into phonological encoding (Dell et al., 1997), were associated with left anterior/mid MTG. Finally, producing neologisms and nonwords was associated with left SMG and the mid portion of STG. In a cohort of individuals with primary progressive aphasia producing semispontaneous speech (Wilson et al., 2010), atrophy in the mid portion of the left temporal lobe (mainly MTG) was associated with the production of nouns of increasing lexical frequency, a measure tapping into both lexical and phonological stages (Kittredge et al., 2008). In turn, producing phonological paraphasias was associated with atrophy in the mid portion of left STG. In sum, converging with the stroke-aphasia literature reviewed above, conceptual/lexical stages tend to be associated with the mid to posterior portions of the temporal lobe, and MTG in particular, in addition to ITG. This latter region is not often represented in stroke-aphasia cases given the difference in arterial blood supply between STG and MTG on the one hand, and ITG on the other. Once phonological representations become implicated, a tendency is seen for associations with (mid) STG and SMG, which is also suggested by some of the stroke studies we reviewed.

Overall, the literature strongly suggests an important role for the left temporal, rather than frontal, lobe in naming, contrary to a perhaps more popular view of the left temporal lobe as the site for comprehension and the frontal lobe for production. Part of this misconception may have its roots in the fact that producing language is a *motor* function, which nevertheless requires the retrieval of conceptual, lexical, and phonological information, processes that we would argue are not particularly linked to the frontal cortex.

Admittedly, the vast majority of the studies reviewed here employed noun rather than verb naming (with the exception of Akinina et al., 2019; Lukic et al., 2021), a deliberate choice to increase the comparability across papers. A review of a large body of literature comparing nouns and verbs has concluded that these two grammatical classes are processed by a largely overlapping set of areas in production and comprehension (Vigliocco et al., 2011). Neural differences between the two emerge, however, as a function of the task (among a few other

variables, see Vigliocco et al., 2011): for example, when there is an emphasis in morphosyntax as in verb naming, in which case the involvement of left IFG becomes more prominent. In both Akinina et al. (2019) and Lukic et al. (2021), the accuracy score was based on any correct morphological form of the verb, thus emphasizing more morphosyntax than noun/object naming does. Thus, one could argue that our inclusion of more studies examining noun naming rather than verb naming may have overemphasized the temporal rather than the frontal cortex. However, the frontal cortex contribution in this case would arguably not be due to conceptual-lexical (and phonological) information being retrieved. Thus, our conclusion remains that the left temporal, rather than the frontal cortex, is the most critical for conceptually driven naming.

The terminology adopted by authors for subdividing the temporal lobe is not always consistent or transparent. Here, we opted for adopting a coordinate-based standard used by Indefrey and Levelt (2004), separating the temporal lobe into an anterior, a mid, and a posterior portion. This system is useful for being more objective (when coordinates are available). However, the boundaries between these portions are not necessarily a subdivision reflecting some form of organization of the temporal lobe (a "natural kind," cf. cytoarchitectonics, transcriptomics, etc.), but rather a convention adopted to help researchers structure results. Future research could attempt to link findings of particular locations in the temporal lobe to physiologically and biologically based subdivisions, which could prove useful for integrating findings across studies using different methods and elucidating temporal lobe functions.

A limitation of our study is the small number of comparable papers in the analyses performed. We used strict criteria whereby papers were excluded if there was cohort overlap or if we could not ascertain that there was no overlap. Hence, more empirical studies are required to increase robustness of the ALE meta-analysis and to strengthen our claims on the psychological nature of the foci identified in the present study. These limitations relate to two recommendations we can make to improve the field of language production. Firstly, authors would ideally disclose cohort overlap with previous studies (this was done in some, but not all, of the studies we reviewed). Secondly, following open science practices, authors would ideally make postprocessed data available (e.g., Thye & Mirman, 2022). An alternative to this solution would be to report a set of coordinates for as much as the method affords. A second limitation of our study is the use of the ALE method without a validation for the probability distribution we employed in combination with LSM data.

In conclusion, the ALE meta-analysis of 10 lesion-symptom mapping studies of naming performance yielded distinct clusters, predominantly in anterior and posterior portions of the left temporal lobe, for which the posterior distribution seems to follow the arcuate fasciculus. Two additional clusters were found in postcentral and middle frontal gyrus. No peaks were identified in the left IFG. Regions consistent with these foci were also revealed by examining papers studying more detailed measures of naming or other populations than stroke, where we found a tendency for lesions in mid to posterior parts of the temporal lobe to be more consistently associated with conceptual-lexical deficits. A major limitation of the present study remains the small number of papers included in the meta-analysis.

AUTHOR CONTRIBUTIONS

Vitória Piai: Conceptualization; Formal analysis; Project administration; Visualization; Writing – original draft; Writing – review & editing. **Dilys Eikelboom**: Formal analysis; Project administration; Visualization; Writing – original draft; Writing – review & editing.

DATA AVAILABILITY STATEMENT

All data associated with these analyses are available via https://osf.io/8xtp9/.

REFERENCES

Akinina, Y., Dragoy, O., Ivanova, M. V., Iskra, E. V., Soloukhina, O. A., Petryshevsky, A. G., Fedina, O. N., Turken, A. U., Shklovsky, V. M., & Dronkers, N. F. (2019). Grey and white matter substrates of action naming. *Neuropsychologia*, *131*, 249–265. https://doi.org/10.1016/j.neuropsychologia.2019.05.015, PubMed: 31129278

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2018a). The behavioural patterns and neural correlates of concrete and abstract verb processing in aphasia: A novel verb semantic battery. *NeuroImage: Clinical*, *17*, 811–825. https://doi.org/10.1016/j.nicl.2017.12.009, PubMed: 29619318

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2018b). Noun and verb processing in aphasia: Behavioural profiles and neural correlates. *NeuroImage: Clinical*, *18*, 215–230. https://doi.org/10.1016/j.nicl.2018.01.023, PubMed: 29868446

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2020a). Mapping psycholinguistic features to the neuropsychological and lesion profiles in aphasia. *Cortex*, *124*, 260–273. https://doi.org/10.1016/j.cortex.2019.12.002, PubMed: 31958653

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2020b). A unified model of post-stroke language deficits including discourse production and their neural correlates. *Brain*, *143*(5), 1541–1554. https://doi.org/10.1093/brain/awaa074, PubMed: 32330940

American Psychological Association. (2023). *APA PsycINFO* (database). https://www.apa.org/pubs/databases/psycinfo/

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry—The methods. *NeuroImage*, *11*(6), 805–821. https://doi.org/10.1006/nimg.2000.0582, PubMed: 10860804

Baldo, J. V., Arévalo, A., Patterson, J. P., & Dronkers, N. F. (2013). Grey and white matter correlates of picture naming: Evidence from a voxel-based lesion analysis of the Boston Naming Test. *Cortex*, *49*(3), 658–667. https://doi.org/10.1016/j.cortex.2012.03.001, PubMed: 22482693

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature Neuroscience*, *6*(5), 448–450. https://doi.org/10.1038/nn1050, PubMed: 12704393

Butler, R. A., Lambon Ralph, M. A., & Woollams, A. M. (2014). Capturing multidimensionality in stroke aphasia: Mapping principal behavioural components to neural structures. *Brain*, *137*(12), 3248–3266. https://doi.org/10.1093/brain/awu286, PubMed: 25348632

Catani, M., & Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, *44*(8), 1105–1132. https://doi.org/10.1016/j.cortex.2008.05.004, PubMed: 18619589

Chen, Q., Middleton, E., & Mirman, D. (2019). Words fail: Lesion-symptom mapping of errors of omission in post-stroke aphasia. *Journal of Neuropsychology*, *13*(2), 183–197. https://doi.org/10.1111/jnp.12148, PubMed: 29411521

Clarivate. (2023). *Web of Science Core Collection* (database). https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/web-of-science/web-of-science-core-collection/

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283, PubMed: 3749399

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*(1–3), 287–314. https://doi.org/10.1016/0010-0277(92)90046-K, PubMed: 1582160

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801, PubMed: 9337631

Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Branch Coslett, H. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, *128*(3), 380–396. https://doi.org/10.1016/j.cognition.2013.05.007, PubMed: 23765000

Devenyi, G. A., Pipitone, J., & Raihaan, P. (2017). *AAL atlas* (Software). https://github.com/CobraLab/documentation/wiki/AAL-Atlas

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, *59*(3), 2349–2361. https://doi.org/10.1016/j.neuroimage.2011.09.017, PubMed: 21963913

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, *30*(9), 2907–2926. https://doi.org/10.1002/hbm.20718, PubMed: 19172646

Faroqi-Shah, Y., Kling, T., Solomon, J., Liu, S., Park, G., & Braun, A. (2014). Lesion analysis of language production deficits in aphasia. *Aphasiology*, *28*(3), 258–277. https://doi.org/10.1080/02687038.2013.853023

Faulkner, J. W., & Wilshire, C. E. (2020). Mapping eloquent cortex: A voxel-based lesion-symptom mapping study of core speech production capacities in brain tumour patients. *Brain and Language*, *200*, Article 104710. https://doi.org/10.1016/j.bandl.2019.104710, PubMed: 31739187

Fridriksson, J., den Ouden, D.-B., Hillis, A. E., Hickok, G., Rorden, C., Basilakos, A., Yourganov, G., & Bonilha, L. (2018). Anatomy of aphasia revisited. *Brain*, *141*(3), 848–862. https://doi.org/10.1093/brain/awx363, PubMed: 29360947

Fridriksson, J., Yourganov, G., Bonilha, L., Basilakos, A., den Ouden, D.-B., & Rorden, C. (2016). Revealing the dual streams of speech processing. *Proceedings of the National Academy of Sciences*, *113*(52), 15108–15113. https://doi.org/10.1073/pnas.1614038114, PubMed: 27956600

Geva, S., Baron, J.-C., Jones, P. S., Price, C. J., & Warburton, E. A. (2012). A comparison of VLSM and VBM in a cohort of patients with post-stroke aphasia. *NeuroImage: Clinical*, *1*(1), 37–47. https://doi.org/10.1016/j.nicl.2012.08.003, PubMed: 24179735

Griffis, J. C., Nenert, R., Allendorfer, J. B., & Szaflarski, J. P. (2017). Damage to white matter bottlenecks contributes to language impairments after left hemispheric stroke. *NeuroImage: Clinical*, *14*, 552–565. https://doi.org/10.1016/j.nicl.2017.02.019, PubMed: 28337410

Halai, A. D., Woollams, A. M., & Lambon Ralph, M. A. (2017). Using principal component analysis to capture individual differences within a unified neuropsychological model of chronic post-stroke aphasia: Revealing the unique neural correlates of speech fluency, phonology and semantics. *Cortex*, *86*, 275–289. https://doi.org/10.1016/j.cortex.2016.04.016, PubMed: 27216359

Halai, A. D., Woollams, A. M., & Lambon Ralph, M. A. (2018a). Predicting the pattern and severity of chronic post-stroke language deficits from functionally-partitioned structural lesions. *NeuroImage: Clinical*, *19*, 1–13. https://doi.org/10.1016/j.nicl.2018.03.011, PubMed: 30038893

Halai, A. D., Woollams, A. M., & Lambon Ralph, M. A. (2018b). Triangulation of language-cognitive impairments, naming errors and their neural bases post-stroke. *NeuroImage: Clinical*, *17*, 465–473. https://doi.org/10.1016/j.nicl.2017.10.037, PubMed: 29159059

Harvey, D. Y., & Schnur, T. T. (2015). Distinct loci of lexical and semantic access deficits in aphasia: Evidence from voxel-based lesion-symptom mapping and diffusion tensor imaging. *Cortex*, *67*, 37–58. https://doi.org/10.1016/j.cortex.2015.03.004, PubMed: 25880795

Hernandez-Pavon, J. C., Makela, N., Lehtinen, H., Lioumis, P., & Makela, J. P. (2014). Effects of navigated TMS on object and action naming. *Frontiers in Human Neuroscience*, *8*, Article 660. https://doi.org/10.3389/fnhum.2014.00660, PubMed: 25228868

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113, PubMed: 17431404

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001, PubMed: 15037128

Isella, V., Rosazza, C., Gazzotti, M., Sala, J., Morzenti, S., Crivellaro, C., Appollonio, I. M., Ferrarese, C., & Luzzatti, C. (2020). A metabolic imaging study of lexical and phonological naming errors in Alzheimer disease. *American Journal of Alzheimer's Disease & Other Dementias*, *35*, Article 1533317520922390. https://doi.org/10.1177/1533317520922390, PubMed: 32356456

Ivanova, M. V., Herron, T. J., Dronkers, N. F., & Baldo, J. V. (2021). An empirical comparison of univariate versus multivariate methods for the analysis of brain–behavior mapping. *Human Brain Mapping*, *42*(4), 1070–1101. https://doi.org/10.1002/hbm.25278, PubMed: 33216425

Kertesz, A. (1982). *The Western aphasia battery*. Grune & Stratton.

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, *25*(4), 463–492. https://doi.org/10.1080/02643290701674851, PubMed: 18704797

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776, PubMed: 11301520

Liljeström, M., Hultén, A., Parkkonen, L., & Salmelin, R. (2009). Comparing MEG and fMRI views to naming actions and objects. *Human Brain Mapping*, *30*(6), 1845–1856. https://doi.org/10.1002/hbm.20785, PubMed: 19378277

Lukic, S., Barbieri, E., Wang, X., Caplan, D., Kiran, S., Rapp, B., Parrish, T. B., & Thompson, C. K. (2017). Right hemisphere grey matter volume and language functions in stroke aphasia. *Neural Plasticity*, *2017*, Article 5601509. https://doi.org/10.1155/2017/5601509, PubMed: 28573050

Lukic, S., Thompson, C. K., Barbieri, E., Chiappetta, B., Bonakdarpour, B., Kiran, S., Rapp, B., Parrish, T. B., & Caplan, D. (2021). Common and distinct neural substrates of sentence production and comprehension. *NeuroImage*, *224*, Article 117374. https://doi.org/10.1016/j.neuroimage.2020.117374, PubMed: 32949711

Mirman, D., Chen, Q., Zhang, Y., Wang, Z., Faseyitan, O. K., Coslett, H. B., & Schwartz, M. F. (2015). Neural organization of spoken language revealed by lesion-symptom mapping. *Nature Communications*, *6*, Article 6762. https://doi.org/10.1038/ncomms7762, PubMed: 25879574

Mirman, D., & Graziano, K. M. (2013). The neural basis of inhibitory effects of semantic and phonological neighbors in spoken word production. *Journal of Cognitive Neuroscience*, *25*(9), 1504–1516. https://doi.org/10.1162/jocn_a_00408, PubMed: 23647518

Mirman, D., & Thye, M. (2018). Uncovering the neuroanatomy of core language systems using lesion-symptom mapping. *Current Directions in Psychological Science*, *27*(6), 455–461. https://doi.org/10.1177/0963721418787486

Mirman, D., Zhang, Y., Wang, Z., Coslett, H. B., & Schwartz, M. F. (2015). The ins and outs of meaning: Behavioral and neuroanatomical dissociation of semantically-driven word retrieval and multimodal semantic recognition in aphasia. *Neuropsychologia*, *76*, 208–219. https://doi.org/10.1016/j.neuropsychologia.2015.02.014, PubMed: 25681739

Na, Y., Jung, J., Tench, C. R., Auer, D. P., & Pyun, S.-B. (2022). Language systems from lesion-symptom mapping in aphasia: A meta-analysis of voxel-based lesion mapping studies. *NeuroImage: Clinical*, *35*, Article 103038. https://doi.org/10.1016/j.nicl.2022.103038, PubMed: 35569227

Pillay, S. B., Binder, J. R., Humphries, C., Gross, W. L., & Book, D. S. (2017). Lesion localization of speech comprehension deficits in chronic aphasia. *Neurology*, *88*(10), 970–975. https://doi.org/10.1212/WNL.0000000000003683, PubMed: 28179469

Piras, F., & Marangolo, P. (2007). Noun–verb naming in aphasia: A voxel-based lesion-symptom mapping study. *NeuroReport*, *18*(14), 1455–1458. https://doi.org/10.1097/WNR.0b013e3282ef6fc9, PubMed: 17712273

Piras, F., & Marangolo, P. (2010). When "Crack walnuts" lies in different brain regions: Evidence from a voxel-based lesion-symptom mapping study. *Journal of the International Neuropsychological Society*, 16(3), 433–442. https://doi.org/10.1017/S1355617710000068, PubMed: 20178682

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062, PubMed: 22584224

Price, C. J., Moore, C. J., Humphreys, G. W., Frackowiak, R. S. J., & Friston, K. J. (1996). The neural regions sustaining object recognition and naming. *Proceedings of the Royal Society B: Biological Sciences*, 263(1376), 1501–1507. https://doi.org/10.1098/rspb.1996.0219, PubMed: 8952093

Pustina, D., Coslett, H. B., Turkeltaub, P. E., Tustison, N., Schwartz, M. F., & Avants, B. (2016). Automated segmentation of chronic stroke lesions using LINDA: Lesion identification with neighborhood data analysis. *Human Brain Mapping*, 37(4), 1405–1421. https://doi.org/10.1002/hbm.23110, PubMed: 26756101

Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, 59, 33–48. https://doi.org/10.1016/j.cortex.2014.07.001, PubMed: 25128898

Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition*, 172, 59–72. https://doi.org/10.1016/j.cognition.2017.12.007, PubMed: 29232595

Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences*, 106(1), 322–327. https://doi.org/10.1073/pnas.0805874106, PubMed: 19118194

Schwartz, M. F., Faseyitan, O., Kim, J., & Coslett, H. B. (2012). The dorsal stream contribution to phonological retrieval in object naming. *Brain*, 135(12), 3799–3814. https://doi.org/10.1093/brain/aws300, PubMed: 23171662

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O. K., Dell, G. S., Mirman, D., & Coslett, H. B. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proceedings of the National Academy of Sciences*, 108(20), 8520–8524. https://doi.org/10.1073/pnas.1014935108, PubMed: 21540329

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Faseyitan, O., Brecher, A., Dell, G. S., & Coslett, H. B. (2009). Anterior temporal involvement in semantic word retrieval: Voxel-based lesion-symptom mapping evidence from aphasia. *Brain*, 132(12), 3411–3427. https://doi.org/10.1093/brain/awp284, PubMed: 19942676

Schwartz, M. F., Wilshire, C. E., Gagnon, D. A., & Polansky, M. (2004). Origins of nonword phonological errors in aphasic picture naming. *Cognitive Neuropsychology*, 21(2–4), 159–186. https://doi.org/10.1080/02643290342000519, PubMed: 21038198

Skipper-Kallal, L. M., Lacey, E. H., Xing, S., & Turkeltaub, P. E. (2017). Functional activation independently contributes to naming ability and relates to lesion site in post-stroke aphasia. *Human Brain Mapping*, 38(4), 2051–2066. https://doi.org/10.1002/hbm.23504, PubMed: 28083891

Stark, B. C., Basilakos, A., Hickok, G., Rorden, C., Bonilha, L., & Fridriksson, J. (2019). Neural organization of speech production: A lesion-based study of error patterns in connected speech. *Cortex*, 117, 228–246. https://doi.org/10.1016/j.cortex.2019.02.029, PubMed: 31005024

Sul, B., Lee, K. B., Hong, B. Y., Kim, J. S., Kim, J., Hwang, W. S., & Lim, S. H. (2019). Association of lesion location with long-term recovery in post-stroke aphasia and language deficits. *Frontiers in Neurology*, 10, Article 776. https://doi.org/10.3389/fneur.2019.00776, PubMed: 31396146

Thye, M., & Mirman, D. (2018). Relative contributions of lesion location and lesion size to predictions of varied language deficits in post-stroke aphasia. *NeuroImage: Clinical*, 20, 1129–1138. https://doi.org/10.1016/j.nicl.2018.10.017, PubMed: 30380520

Thye, M., & Mirman, D. (2022). *Relative contributions of lesion location and lesion size to predictions of varied language deficits in post-stroke aphasia* [Project]. https://osf.io/w4u2q/

Tochadse, M., Halai, A. D., Lambon Ralph, M. A., & Abel, S. (2018). Unification of behavioural, computational and neural accounts of word production errors in post-stroke aphasia. *NeuroImage: Clinical*, 18, 952–962. https://doi.org/10.1016/j.nicl.2018.03.031, PubMed: 29876280

Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage*, 16(3), 765–780. https://doi.org/10.1006/nimg.2002.1131, PubMed: 12169260

Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human Brain Mapping*, 33(1), 1–13. https://doi.org/10.1002/hbm.21186, PubMed: 21305667

Tyler, L. K., Marslen-Wilson, W., & Stamatakis, E. A. (2005). Dissociating neuro-cognitive component processes: Voxel-based correlational methodology. *Neuropsychologia*, 43(5), 771–778. https://doi.org/10.1016/j.neuropsychologia.2004.07.020, PubMed: 15721189

Urgesi, C., Candidi, M., & Avenanti, A. (2014). Neuroanatomical substrates of action perception and understanding: An anatomic likelihood estimation meta-analysis of lesion-symptom mapping studies in brain injured patients. *Frontiers in Human Neuroscience*, 8, Article 344. https://doi.org/10.3389/fnhum.2014.00344, PubMed: 24910603

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews*, 35(3), 407–426. https://doi.org/10.1016/j.neubiorev.2010.04.007, PubMed: 20451552

Walker, G. M., Schwartz, M. F., Kimberg, D. Y., Faseyitan, O., Brecher, A., Dell, G. S., & Coslett, H. B. (2011). Support for anterior temporal involvement in semantic error production in aphasia: New evidence from VLSM. *Brain and Language*, 117(3), 110–122. https://doi.org/10.1016/j.bandl.2010.09.008, PubMed: 20961612

Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., Miller, B. L., & Gorno-Tempini, M. L. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain*, 133(7), 2069–2088. https://doi.org/10.1093/brain/awq129, PubMed: 20542982

Zhang, Y., Kimberg, D. Y., Coslett, H. B., Schwartz, M. F., & Wang, Z. (2014). Multivariate lesion-symptom mapping using support vector regression. *Human Brain Mapping*, 35(12), 5861–5876. https://doi.org/10.1002/hbm.22590, PubMed: 25044213

Zhao, Y., Halai, A. D., & Lambon Ralph, M. A. (2020). Evaluating the granularity and statistical structure of lesions and behaviour in post-stroke aphasia. *Brain Communications*, 2(2), Article fcaa062. https://doi.org/10.1093/braincomms/fcaa062, PubMed: 32954319