The Neural Basis of Reversible Sentence Comprehension: Evidence from Voxel-based Lesion Symptom Mapping in Aphasia

Malathi Thothathiri[1,2]*, Daniel Y. Kimberg[2], and Myrna F. Schwartz[1]

Abstract

■ We explored the neural basis of reversible sentence comprehension in a large group of aphasic patients ($n = 79$). Voxel-based lesion symptom mapping revealed a significant association between damage in temporo-parietal cortex and impaired sentence comprehension. This association remained after we controlled for phonological working memory. We hypothesize that this region plays an important role in the thematic or what–where processing of sentences. In contrast, we detected weak or no association between reversible sentence comprehension and the ventrolateral pFC, which includes Broca's area, even for syntactically complex sentences. This casts doubt on theories that presuppose a critical role for this region in syntactic computations. ■

INTRODUCTION

Neuroimaging and lesion studies have consistently implicated both anterior and posterior peri-sylvian areas in sentence comprehension, but the separate functional contributions of these areas still need to be clarified. Sentence comprehension is a complex process involving access to different linguistic representations; manipulation of these representations in memory; and other possibly nonlinguistic, task-sensitive processes such as cognitive control. This complexity leads to difficulty in determining the functional basis of comprehension deficits after brain damage and activation during neuroimaging studies. In this study, we attempted to clarify the role of frontal and temporo-parietal regions in sentence comprehension using voxel-based lesion symptom mapping (VLSM; Bates et al., 2003) in a large cohort of aphasic patients.

Many neuroimaging studies have found greater activation in ventrolateral pFC (VLPFC) for complex sentences compared with their simpler counterparts (Fiebach, Schlesewsky, Lohmann, von Cramon, & Friederici, 2005; Ben-Shachar, Hendler, Kahn, Ben-Bashat, & Grodzinsky, 2003; Caplan, Alpert, & Waters, 1998, 1999; Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Stromswold, Caplan, Alpert, & Rauch, 1996). Broadly speaking, this evidence has been interpreted in two different ways. Syntax-oriented theories propose that VLPFC—particularly BA 44 and BA 45, referred to jointly as Broca's area—supports sentence comprehension via core syntactic computations (e.g., Ben-Shachar et al., 2003; Caplan et al., 1998). As such, VLPFC involvement is necessary for accurate comprehension, especially for more complex sentences. In contrast, resource-oriented theories propose that VLPFC supports sentence comprehension via processes such as working memory and cognitive control (Fiebach et al., 2005; Novick, Trueswell, & Thompson-Schill, 2005). These accounts are consistent with the view that VLPFC plays an indirect role in comprehension via its modulatory function over linguistic representations that reside *elsewhere*. The need for such modulatory function would depend upon both stimulus and task factors. For example, VLPFC involvement may be most required when a prepotent interpretation must be overridden in favor of a less preferred meaning (Novick et al., 2005). At stake is the accurate characterization of the function of VLPFC, particularly Broca's area, in language. Here we analyzed the relation between VLPFC damage and comprehension deficits in aphasia to elucidate whether this region plays a necessary role in sentence comprehension.

Posterior peri-sylvian areas including superior and middle temporal gyri and the inferior parietal lobule are thought to underlie many component processes of sentence comprehension, including storage of semantic representations, verbal working memory, and verb argument structure processing (den Ouden, Fix, Parrish, & Thompson, 2009; Lau, Phillips, & Poeppel, 2008; Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004; Hickok & Poeppel, 2004; Wise, 2003; Caplan et al., 2001). Here we aimed to identify the specific temporal and parietal areas that are critical for sentence comprehension. Furthermore, we controlled for covariates of sentence comprehension ability, such as phonological working memory, and asked whether the role of temporo-parietal areas in sentence comprehension extends beyond phonological and lexical level processes.

[1]Moss Rehabilitation Research Institute, Philadelphia, PA, [2]University of Pennsylvania
*Now at Swarthmore College.

We addressed these questions by performing VLSM analyses on sentence comprehension scores from 79 left-hemisphere stroke patients with aphasia. Sentences were semantically reversible (e.g., The man served the woman) such that accurate comprehension could not be achieved by using semantic knowledge of who typically does what. Such reversible sentences are often used in psycholinguistic and neurolinguistic studies to isolate "syntactic" comprehension processes, that is, the derivation of the correct meaning of a sentence from its syntactic form.

VLSM is a technique used to investigate voxel-wise correlations between lesion status and behavior. A test statistic (e.g., t value) is computed based on the difference in behavioral score between participants who have damage to a particular voxel and those who do not. Effects across the whole brain are then evaluated for significance after correcting for multiple comparisons. Differences between VLSM and traditional methods of overlaying lesions have been discussed extensively elsewhere (Kimberg, Coslett, & Schwartz, 2007; Rorden & Karnath, 2004; Bates et al., 2003). Among other things, VLSM uses inferential and not merely descriptive statistics; it can be applied to continuous (i.e., nondichotomized) behavioral data from a variety of patients with different behavioral and lesion profiles, and it avoids having to choose predetermined ROIs. Of these features, the last is less clearly an advantage for the current study, because previous studies of sentence processing provide strong rationale for focusing on particular anterior and posterior brain regions. With that in mind and to see whether evidence from regional analyses would converge with the VLSM results, we additionally computed the association between behavioral scores and lesion extent in several a priori Brodmann's areas.

METHODS

Subjects

We analyzed data from 79 patients recruited from the patient registry at the Moss Rehabilitation Research Institute (Schwartz, Brecher, Whyte, & Klein, 2005). All patients had aphasia following left hemisphere stroke and met the following criteria: English as primary language, adequate vision and hearing, premorbid right-handedness, no major psychiatric or neurologic comorbidities, and a left-hemisphere cortical lesion confirmed via MRI or CT. All patients scored 80% or higher on a lexical comprehension test (see below), allowing us to evaluate variability in syntactic comprehension independent of lexical deficits. The patient group was diverse. On the Western Aphasia Battery (Kertesz, 1982), the aphasia quotient range was 38–97.6 (mean = 78.8), and the aphasia subtype breakdown was 34 anomic, 21 Broca's, 15 conduction, 1 transcortical motor, and 8 "recovered" with aphasia quotient of >93.8.

Patients gave written permission to participate under a protocol approved by the Albert Einstein Medical Center Institutional Review Board. Structural images were obtained under a protocol at the University of Pennsylvania School of Medicine.

Sentence Comprehension Behavioral Measure

Sentence comprehension scores were obtained from a two alternative forced-choice sentence-to-picture matching task, which is part of the Philadelphia Comprehension Battery (Saffran, Schwartz, Linebarger, Martin, & Bochetto, 1988). Patients were tested on 30 sentences, five each of actives, actives with prepositional phrases,[1] passives, locatives, subject relative clauses, and object relative clauses.

For each sentence type, the lexical content in the target and distractor pictures was the same. Participants could succeed only by assigning thematic roles in accordance with the syntax. For example, for the sentence "The man serves the woman," one picture showed a man serving a woman and the other showed a woman serving a man (see Table 1 for example stimuli). Our primary behavioral measure was percent correct averaged across all sentence types (SentComp). We also analyzed percent correct on a subset of sentence types where necessary.

Lexical Comprehension Screen

We only included those patients who scored 80% or higher on a lexical version of the sentence comprehension test. These "lexical" trials were interspersed with the critical

Table 1. Example Stimuli in Sentence Comprehension Test

Type	Sentence	Target Picture	Distractor Picture
Active	The girl washes the boy.	girl washing boy	boy washing girl
Active-prep	The policeman shoots the robber in the alley.	policeman shooting robber	robber shooting policeman
Loc	The plate is under the napkin.	plate under napkin	napkin under plate
Passive	The man is served by the woman.	woman serving man	man serving woman
Subj-rel	The dog that followed the hunter was alert.	dog following hunter	hunter following dog
Obj-rel	The girl that the boy washed was talkative.	boy washing girl	girl washing boy

Active-prep: actives with prepositional phrases; Loc: locatives; Subj-rel: subject relative clauses; Obj-rel: object relative clauses.

trials described above and followed the same procedure. For each sentence type, the lexical trials could be answered by paying attention to lexical content alone. For example, for the sentence "The dog chases the boy," the two accompanying pictures were of a dog chasing a boy (target) and of a dog chasing a rabbit (distractor). By including only patients who passed the lexical screen, we could be confident that the variance in the sentence comprehension scores was because of syntactic and not lexical comprehension.

Other Behavioral Measures

In secondary analyses, we controlled for phonological working memory as measured by rhyme probe span[2] and the nonword repetition test (NWRep). The rhyme probe span measures the list length for which participants can correctly judge whether a probe word rhymes with one of the items in the preceding list. Participants listened to a list and an immediately following probe word and responded "yes" or "no" (e.g., List: black–more–some, Probe: plum, Response: yes; List: fright–chill–threw, Probe: steep, Response: no). The test starts with List Length 1 and advances to the next length if the participant scores at least 75% correct. The maximum score is 9. In the NWRep, participants were asked to repeat single nonwords of one, two, or three syllables (e.g., fos, tayson, and dunapour, respectively) immediately after hearing them. The score is a simple percent correct out of 60 items.

VLSM Analysis

We analyzed structural images acquired using MRI ($n = 43$) or CT ($n = 36$). Details of imaging, segmentation, and registration procedures have previously been reported (Schwartz et al., 2009). For patients with MRI scans, lesions were first drawn manually on a $1 \times 1 \times 1$ mm T1-weighted structural image. Before warping, lesions were masked, and the structural scans and lesion maps were registered to a common template constructed from images acquired on the same scanner using a symmetric diffeomorphic registration algorithm (Avants, Schoenemann, & Gee, 2006; see also www.picsl.upenn.edu/ANTS/). A single mapping from this intermediate template to the 1-mm Montreal Neurological Institute (MNI) space "Colin27" volume (Holmes et al., 1998) was used to complete the mapping from subject space to MNI space. The final lesion map was quantized to produce a 0/1 map using 0.5 as the cutoff. An experienced neurologist who was blind to the behavioral data verified depictions of the lesions. For patients with CT scans, the same neurologist drew lesion maps directly onto the Colin27 volume.

We excluded voxels in which fewer than five patients had lesions.[3] All analyses were done using VoxBo (www.voxbo.org). We computed t statistics comparing patients with and without lesions in each voxel. The resulting t map was thresholded to control for false discovery rate (FDR;

Genovese, Lazar, & Nichols, 2002) at $q = 0.01$, wherein q is the expected proportion of false positives among suprathreshold voxels. Although our 0.01 threshold is the most stringent of FDR thresholds used in the literature, it should be borne in mind that the FDR correction is relatively relaxed, in that it controls for expected proportion of false positives among suprathreshold voxels and not for likelihood of any false positives, as is the case with Bonferroni correction or other methods for control of the family-wise error rate (Kimberg et al., 2007).

Regional Analysis

As noted, we supplemented the voxel-based analysis with regional analyses computing the association between behavioral scores and percent damage in 10 a priori Brodmann's areas (BA 7, BA 20, BA 21, BA 22, BA 37, BA 38, BA 39, BA 40, BA 44, and BA 45). This list, although nonexhaustive, includes areas frequently implicated in language processing, such as those in the inferior frontal region (BA 44 and BA 45); inferior, middle, and anterior temporal region (BA 20, BA 21, BA 22, BA 37, and BA 38); the junction of temporal and parietal cortices (BA 39 and BA 40); and the superior parietal lobe (BA 7). These percent damage scores were obtained using a modified left-hemisphere-only version of the Brodmann's atlas available with MRIcron. Each lesion mask was overlaid on the atlas, and using a feature in VoxBo, the proportion of overlapping voxels between the lesion mask and our ROI was calculated.

The two methods—VLSM and regional analysis—differ in spatial scale and have complementary strengths and weaknesses. Regional analysis has reduced power to detect patterns that are only present in a subset of voxels within the region. The advantage that VLSM holds in this regard is one reason why we chose to use it in the main analysis. On the other hand, VLSM requires a much greater number of statistical tests and, therefore, incurs a steeper correction for multiple comparisons to maintain the same overall false positive rate. Thus, associations with sentence comprehension that are represented among many temporo-parietal or frontal voxels but are weak in magnitude could be missed by VLSM but detected by regional analysis. In the ideal case, negative findings obtained with VLSM for temporo-parietal or frontal areas would be confirmed by the regional analysis, lending confidence that these were not method-dependent Type 2 errors. Convergence onto the same positive findings would assuage concerns about the false positive rate tolerated by the FDR correction method.

RESULTS

Behavioral Measures

Results from the behavioral tests are shown in Table 2. Sentence comprehension accuracy ranged between 37% and 100%, with a mean of 74.8%. Scores grouped by subsets

Table 2. Behavioral Scores

Test/Measure	Mean (SD)	Median	Min	Max
SentComp-All (% correct)	74.8 (16.5)	73.3	37	100
SentComp-Canonical (% correct)	84.6 (18.1)	90	20	100
SentComp-Noncanonical (% correct)	62.2 (24.2)	60	10	100
SentComp-1 Proposition (% correct)	78.2 (19.8)	80	30	100
SentComp-2 Proposition (% correct)	68.5 (21.3)	70	20	100
Rhyme probe span (max. list length)	3.2 (1.7)	3	0.5	7.3
NWRep[a] score (% correct)	53 (25.7)	53	0	98

[a]Two patients did not complete this test. Statistics are computed from 77 patients.

of sentence types showed the expected effects of canonicity and the number of propositions. In canonical English structures, the syntactic subject is the agent of the action being described; in noncanonical structures, the subject is the patient. Under many linguistic theories, noncanonical structures are derived by the movement of a syntactic constituent from its canonical structural position to a different one (e.g., Grodzinsky & Santi, 2008). Thus, noncanonical structures are often harder to process than their canonical counterparts by virtue of their being less frequent and/or more syntactically complex. Performance in our task was consistent with this general pattern: accuracy on noncanonical sentence types (passives, object relatives) was significantly lower than that on canonical sentence types (actives, subject relatives; mean = 62.2% vs. 84.6%, $F(1, 78) = 82.2, p < .001$).[4]

We also split sentences according to the number of propositions. Two-proposition sentences (subject relatives, object relatives) require maintaining and matching multiple propositions to the corresponding picture and could consequently be harder than one-proposition sentences (actives, passives). Consistent with this idea, accuracy in sentence-to-picture matching was significantly lower for the former compared with the latter (mean = 68.5% vs. 78.2%, $F(1, 78) = 21.7, p < .001$).

Anatomical Findings

Comparing Lesions in Frontal and Temporo-parietal Areas

The focus of the current study is to evaluate the relationship between reversible sentence comprehension and damage in frontal and temporo-parietal areas. In VLSM, the power to detect brain–behavior relationships at a given voxel depends on the number of patients in the lesioned and unlesioned groups. An optimal distribution is half and half (in the current data set, ~39 lesioned and unlesioned in any voxel). Figure 1 shows the lesion overlap map for all 79 patients. Coverage was good in both regions. The maximum number of lesions in a BA 44/BA 45 voxel was 37; the maximum for BA 22/BA 39/BA 40 was 35. Thus, a priori power to detect a given effect size should be roughly comparable for voxels in the two regions.

Overall Sentence Comprehension

First, we explored the voxelwise associations between lesion status and comprehension of all sentence types. We entered SentComp scores into the VLSM analysis. Results showed a large suprathreshold cluster in the temporo-parietal region and no suprathreshold voxels in VLPFC
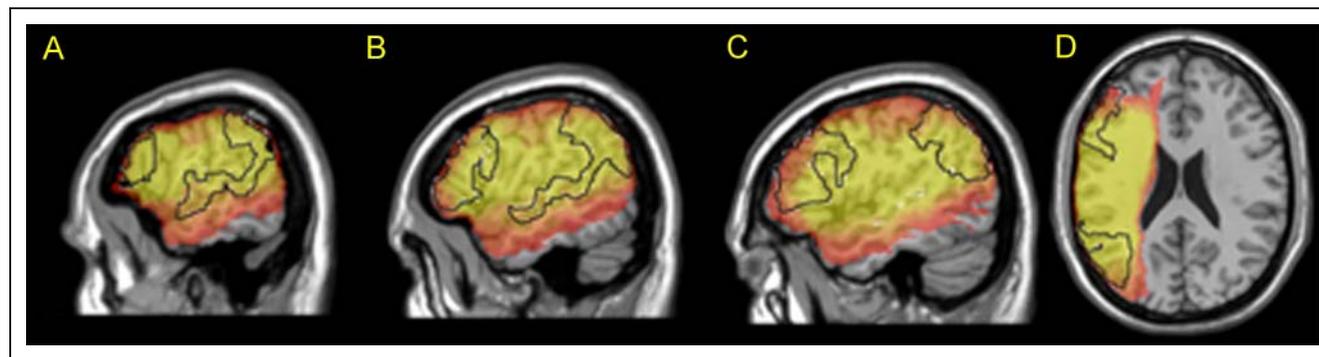
Figure 1. Lesion overlap map for 79 patients. ROIs (BA 44/BA 45 and BA 22/BA 39/BA 40) are indicated by black outlines. Maps A–C are at MNI x coordinates −60, −54, and −48. Map D is a single axial slice at $z = 22$. Number of lesions in each voxel is rendered in a red (5) to yellow (20 and above) scale.
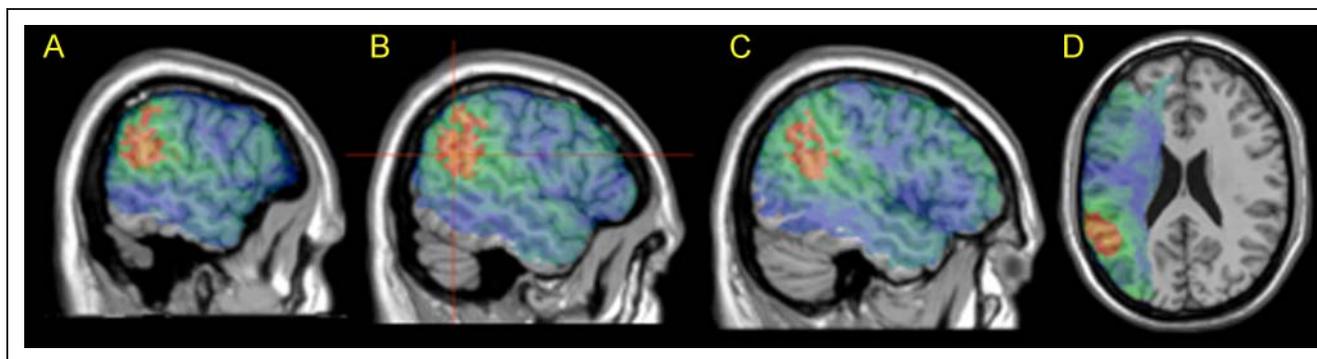
Figure 2. Sentence comprehension scores. Voxels exceeding the FDR threshold ($q = 0.01$) are rendered in a red ($t = 3.61$) to yellow ($t > 5$) scale. Nonsignificant values are rendered in a green (t just below threshold) to blue ($t = 0$ or below) scale. Maps A–C are at MNI x coordinates −60, −54, and −48. Map D is a single axial slice at $z = 22$. The peak t value of 5.42 was centered on MNI coordinates −54, −55, 22 (three voxels), which is indicated by crosshairs in B.

(Figure 2; 12,661 suprathreshold voxels spanning BA 21, BA 22, BA 39, and BA 40).

This finding was corroborated by regional analyses. We explored the regionwise correlations between sentence comprehension scores and percent damage in 10 Brodmann's areas (see Methods). In bivariate analyses, SentComp scores correlated significantly with percent damage in BA 7 ($r = −.25$), BA 21 ($r = −.25$), BA 22 ($r = −.37$), BA 39 ($r = −.43$), and BA 40 ($r = −.41$; all $ps < .05$). There was also a marginal correlation with BA 44 ($r = −.21, p = .06$). Correlations between SentComp and damage in BA 20, BA 37, BA 38, and BA 45 were not significant ($p > .1$). We entered overall lesion volume and percent damage in those areas that marginally or significantly correlated with the behavioral scores (BA 7, BA 21, BA 22, BA 39, BA 40, and BA 44) into a simultaneous regression. The overall model was significant ($R^2 = .29, p < .002$), with marginally significant independent contributions from percent damage in BA 22 ($\beta = −.45, p = .06$) and BA 39 ($\beta = −.26, p = .1$) only. All the other predictors, including lesion volume, were not significant ($p > .1$).

Our results, thus far, lend support to a statistically significant relationship between temporo-parietal regions and sentence comprehension and offer weak or no support for a similar relationship between frontal regions and sentence comprehension. Additionally, we could ask if there is a meaningful *difference* between the effects of temporo-parietal versus frontal lesions on reversible sentence comprehension. We evaluated this question using a post hoc regional analysis. We compared the correlations between comprehension accuracy and percent damage in BA 22/BA 39/BA 40 (combined together) and BA 44/BA 45 (combined together). The correlation in the frontal region was −.206, whereas the correlation in the temporo-parietal region was −.495. The difference between these two correlations was significant by the Hotelling–Williams test ($t_{HW} = 2.19, p < .05$; Steiger, 1980). This suggests a stronger relationship between comprehension accuracy and temporo-parietal compared with frontal lesions.

Effect of Canonicity

Broca's area within VLPFC has been linked in particular to the comprehension of noncanonical sentences. It has been theorized that this might reflect either increased syntactic working memory demands or the specific processing of syntactic movement (Ben-Shachar et al., 2003). To investigate possible differences between canonical and noncanonical sentence comprehension, we performed separate analyses for comprehension scores on actives and subject relatives (canonical) and passives and object relatives (noncanonical). The results are shown in Figure 3. Significant results were found in temporo-parietal voxels for both sentence groups (canonical: 3009 suprathreshold voxels in BA 39 and BA 40; noncanonical: 10,439 suprathreshold voxels in BA 22, BA 39, and BA 40). There were no suprathreshold voxels in VLPFC for either sentence group.

In regional analyses, we examined bivariate correlations between canonical and noncanonical comprehension scores and the 10 chosen Brodmann's areas. Both canonical and
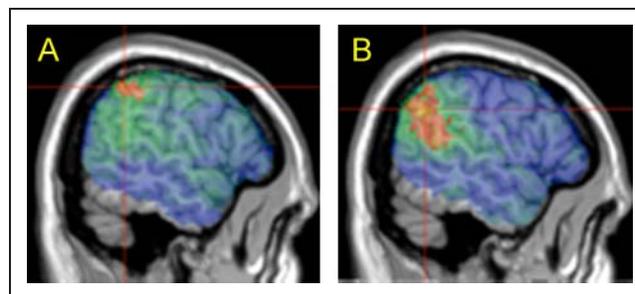
Figure 3. (A) Canonical comprehension scores. Voxels exceeding the FDR threshold ($q = 0.01$) are rendered in a red ($t = 3.96$) to yellow ($t > 5$) scale. The peak t value of 5.31 was centered on MNI coordinates −57, −47, 46 (11 voxels), indicated by crosshairs. (B) Noncanonical comprehension scores. Voxels exceeding the FDR threshold ($q = 0.01$) are rendered in a red ($t = 3.66$) to yellow ($t > 5$) scale. The peak t value of 5.57 was centered on MNI coordinates −56, −54, 32 (8 voxels), indicated by crosshairs.
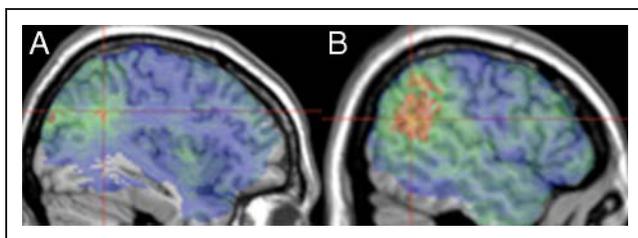
Figure 4. (A) Sentence comprehension scores with rhyme probe spans as covariates. Voxels exceeding the FDR threshold ($q = 0.01$) are rendered in a red ($t = 4$) to yellow ($t > 5$) scale. A peak t value of 5.15 was found at MNI coordinates −38, −53, 31 (one voxel) indicated by crosshairs. (B) Sentence comprehension scores with NWRep scores as covariates. Voxels exceeding the FDR threshold ($q = 0.01$) are rendered in a red ($t = 3.62$) to yellow ($t > 5$) scale. The peak t value of 5.35 was centered at MNI coordinates −53, −55, 22 (three voxels), indicated by crosshairs.

noncanonical scores were significantly correlated with percent damage in BA 7, BA 22, BA 39, and BA 40 (ps < .05). In addition, there was a marginal correlation between canonical scores and damage in BA 44 ($p = .07$).

We also analyzed the difference between noncanonical and canonical scores. For the VLSM analysis, we were unable to obtain a threshold for FDR = 0.01. In the regional analysis, the difference score was significantly correlated with damage in BA 39 ($p < .05$) and no other region.

Phonological Working Memory

What functional interpretation should we assign to the temporo-parietal effects? Areas near the posterior tip of the sylvian fissure, at the junction of temporal and parietal cortices, have been hypothesized to underlie phonological working memory and/or sensori-motor integration (Hickok & Poeppel, 2004; Narain et al., 2003). We evaluated whether the role of temporo-parietal areas in sentence comprehension extends beyond these functions.

To control for phonological working memory, we entered patients' rhyme probe spans as covariates in the VLSM analysis of SentComp scores. This analysis yielded suprathreshold voxels in temporo-parietal areas only (Figure 4A; 3405 voxels in BA 21, BA 22, BA 39, and BA 40).

Repetition tasks are a widely accepted means of measuring sensori-motor integration. We entered patients' NWRep scores as covariates in the VLSM analysis of SentComp scores. Two patients did not complete NWRep, so this analysis was performed on measures from the remaining 77 patients. Figure 4B shows the suprathreshold voxels, which were in temporo-parietal areas only (12,170 voxels in BA 21, BA 22, BA 39, and BA 40).

These results suggest that temporo-parietal regions play an additional role in sentence comprehension that cannot be reduced to phonological working memory.[5] In regional analyses, we entered percent damage in BA 21, BA 22, BA 39, and BA 40 in a simultaneous regression with rhyme probe span. This revealed significant independent contribu-

tions from rhyme probe span ($\beta = .4$) and BA 39 damage ($\beta = -.28$). A similar simultaneous regression with NWRep scores replacing rhyme probe span revealed that NWRep was a significant predictor of SentComp scores ($\beta = .26$) and BA 39 damage made a marginal independent contribution ($\beta = -.24; p = .07$).

Task Demands

Sentence comprehension as measured by the sentence-to-picture matching task involves task-related resources that may be different from core comprehension processes. Participants need to hold the linguistically derived sentence interpretation in memory and compare it with their analyses of the two pictures. The demands of this matching process may be increased when participants have to maintain two propositions as opposed to one proposition in mind. Thus, subject and object relative sentences (e.g., The girl that washed the boy was talkative), which contain two propositions, may recruit more task-related resources than single-proposition active and passive sentences (Waters, Rochon, & Caplan, 1998). In our task, successful selection of the target picture did not require processing of the second proposition (e.g., that the girl was talkative). Nevertheless, we evaluated whether different brain regions would show differential sensitivity to the number of propositions.

We were unable to obtain FDR thresholds for the VLSM analysis of two-proposition minus one-proposition scores (subject + object relatives minus actives + passives). Figure 5, therefore, shows the uncorrected t map ($t > 1.67$) from this analysis. Importantly for our purposes, there was no suggestion of a preferential relation between the number of propositions and temporo-parietal areas. In fact, the highest t values for this analysis were in the frontal
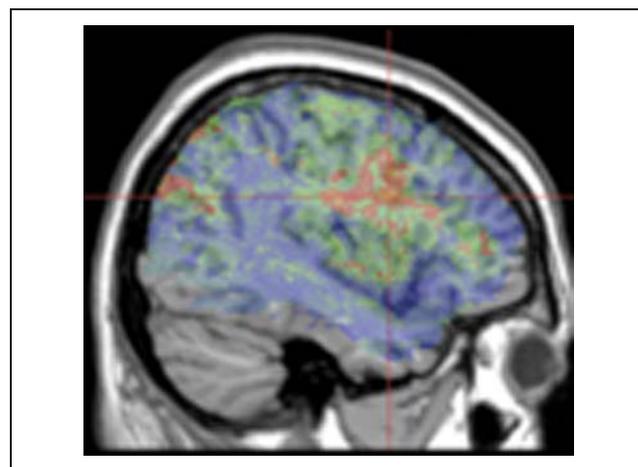
Figure 5. Two-proposition minus one-proposition comprehension scores. Voxels exceeding an *uncorrected t* test threshold are rendered in a red ($t = 1.67$) to yellow ($t > 5$) scale. The peak t value of 4.03 was centered at MNI coordinates −42, 5, 26 (two voxels) indicated by crosshairs.
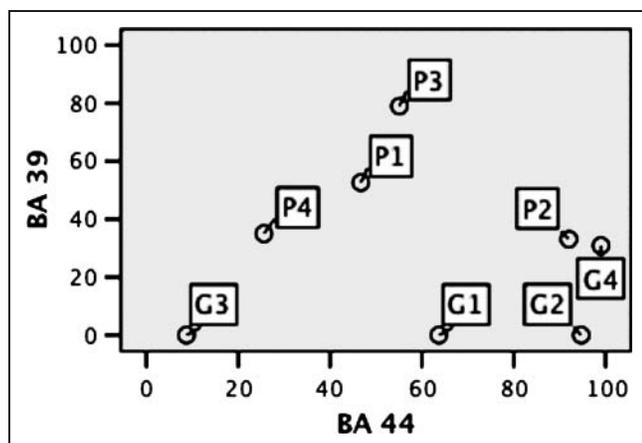
Figure 6. Percent damage in BA 39 and BA 44 for eight Broca's aphasic poor (P) and good (G) comprehenders.

cortex—in BA 44. This is compatible with the interpretation that this region subserves the task-related resources (e.g., working memory) that are additionally required for two-proposition compared with one-proposition sentences. However, any such interpretation would have to be tentative because of the lack of inferential statistics.

Broca's Aphasia

Evidence from Broca's aphasia is often cited in favor of a critical role for Broca's area in syntactic comprehension. The underlying assumptions behind this inference are problematic (see Discussion). In our sample, the comprehension abilities of the 21 Broca's aphasics covered the spectrum: 10 performed poorly (≤65%), 6 performed well above chance (≥80%), and 5 scored somewhere in be-

tween, suggesting no special relationship between Broca's aphasia and impaired syntactic comprehension. Lesion comparisons also throw doubt on the presumed link between impaired comprehension and damage in Broca's area. Figure 6 shows percent damage in BA 39 and BA 44 for two roughly matched groups of Broca's aphasics (four poor and four good comprehenders; Table 3). Most good and poor comprehenders had extensive BA 44 damage (>40%), but the poor comprehenders tended to have more damage in BA 39.

DISCUSSION

The neural basis of comprehension has long been an object of study (Wernicke, 1874). Early studies relied on small patient samples and imprecise neuroanatomical data. State-of-the-art imaging technology now makes it possible to analyze lesion-deficit relations more precisely. Recent studies using functional connectivity analysis and voxel-based morphometry have begun to elucidate how damage or atrophy in anterior and posterior brain regions might impact component processes of sentence comprehension (Warren, Crinion, Ralph, & Wise, 2009; Amici et al., 2007). We used VLSM to address similar questions (see also Dronkers et al., 2004; Bates et al., 2003). We explored the role of VLPFC and temporo-parietal areas in reversible sentence comprehension as evaluated by a sentence-to-picture matching task. Lesion coverage was roughly comparable in the two areas, yet voxels carrying an association between comprehension accuracy and lesion status were identified in temporo-parietal areas but not VLPFC. Results of the regional analyses generally corroborated what we found with VLSM. Correlations between comprehension scores and percent damage for

Table 3. Two Groups of Roughly Matched Broca's Aphasics

Patient Group	Age at Testing	Education (years)	WAB Fluency	Lesion Volume (cc)	SentComp Accuracy (%)
Poor Comprehenders					
MR0047	64	13	4-5	171.6	53
MR0190	59	16	4	205.3	47
MR1737	43	12	4	92.6	63
MR0057	63	16	4	148.8	60
Average	57.3	14.3	4.3	154.6	55.8
Good Comprehenders					
MR1238	49	14	4	172.2	100
MR0046	68	12	4	197.3	93
MR0865	56	12	4	71.8	93
MR1283	54	16	4-5	307.9	80
Average	56.8	13.5	4.3	187.3	91.5

a priori Brodmann's areas were significant for one or more temporo-parietal areas (BA 22/BA 39/BA 40), whereas in frontal regions, they were nonsignificant (BA 45) or marginally significant (BA 44). A post hoc analysis also showed that the correlation between SentComp and percent damage was significantly greater in the temporo-parietal region than in the frontal region.

Temporo-parietal Cortex

Across different analyses, we found a robust association between damage in posterior peri-sylvian areas and impaired reversible sentence comprehension. VLSM results consistently implicated voxels at the junction of temporal and parietal lobes (BA 21/BA 22 and BA 39/BA 40). These results are consistent with other VLSM studies, which have also found an association between auditory comprehension and posterior temporal and inferior parietal areas (Dronkers et al., 2004; Bates et al., 2003). Our regional analyses corroborate the VLSM results: Both overall sentence comprehension scores and separate canonical and noncanonical sentence comprehension scores were significantly correlated with percent damage in BA 22, BA 39, and BA 40. BA 22 and BA 39 continued to show a marginal correlation with sentence comprehension scores even after controlling for lesion volume. As described below, our results extend previous VLSM findings in two other important ways.

First, our task was specifically designed to test the assignment of thematic roles based on syntax. We isolated syntactic comprehension processes by screening participants on a lexical version of our task so that failure in sentence comprehension was unlikely to be because of impaired lexical comprehension. Previous VLSM studies (Dronkers et al., 2004; Bates et al., 2003) sought to identify areas related to all aspects of auditory comprehension. Thus, they used tests like WAB or CYCLE-R (Curtiss & Yamada, 1988) and tested aphasic patients across the spectrum, including those with lexical comprehension deficits. The results from these studies identified the middle temporal gyrus as a critical area for some aspect of auditory comprehension, likely lexical processing (Dronkers et al., 2004). Interestingly, these studies also reported findings in BA 39 consistent with our own results.

We also extend previous findings in showing that the role of temporo-parietal areas in sentence comprehension is not limited to their role in phonological working memory and sensori-motor integration. Dronkers et al. (2004) interpreted their temporo-parietal results in such terms. This is in line with recent proposals regarding the roles of different posterior peri-sylvian areas. Hickok and Poeppel (2004) have proposed that bilateral superior temporal gyrus is responsible for acoustic–phonetic processing, left inferior temporal cortex for the sound–meaning interface, and the boundary between parietal and temporal lobes near the sylvian fissure (sylvian parietal–temporal) for the sensori-motor interface. They propose that the role of the sensori-motor interface is to support verbal working memory; in their view, the ability to use articulatory processes (rehearsal) keep auditory representations active. Similarly, Wise and colleagues have reported several results that implicate the temporo-parietal region in sentence comprehension. They interpret these results as reflecting working memory and/or sensori-motor processing (Crinion, Lambon-Ralph, Warburton, Howard, & Wise, 2003; Narain et al., 2003; Wise, 2003; Scott, Blank, Rosen, & Wise, 2000). The voxels where we found a significant association with comprehension scores include those in the temporo-parietal cortex. These associations remained even after controlling for rhyme probe span and NWRep score. Thus, although we do not argue against a role for temporo-parietal cortex in the above-mentioned working memory and sensori-motor processes, we have evidence that this region plays some other additional role in sentence comprehension.

Our results are consistent with neuroimaging studies that have used different paradigms than sentence-to-picture matching. It is notable that some studies whose main focus was VLPFC have nevertheless found significant results in left posterior temporal and/or inferior parietal cortices (e.g., Ben-Shachar, Palti, & Grodzinsky, 2004; BA 39 bordering BA 22 and BA 37). Two recent neuroimaging studies emphasized the involvement of left temporal and parietal regions to the exclusion of VLPFC. The first study compared German canonical and noncanonical sentences equated for accuracy and RTs (i.e., difficulty) and found increased activation for noncanonical sentences in superior and posterior temporal cortex (BA 21/BA 22 and BA 41/BA 42; Wartenburger et al., 2004). The second study compared Japanese passive and active sentences and found increased activation for the passives in the inferior parietal lobule (BA 39; Yokoyama et al., 2007).

Our results are also consistent with a study that found significant correlations between syntactic comprehension and lesion extent/PET metabolism in temporo-parietal areas in aphasic patients (Caplan et al., 2007). For example, accuracy in their sentence-to-picture matching task was significantly correlated with percent lesion volume in Wernicke's area and PET activity in the inferior parietal lobe. Performance in an object manipulation task was correlated with percent lesion volume in the inferior and superior parietal lobes.

Collective evidence, thus, favors a role for temporo-parietal regions in sentence comprehension. We have already argued that this role may not be restricted to working memory functions. Reversible sentence comprehension requires the computation of thematic relations between sentence constituents (who did what to whom). This might require the dynamic binding of "what" (a particular entity) and "where" (the role that an entity plays in an event; Chang, Dell, & Bock, 2006). The auditory system may contain a dorsal–ventral partitioning similar to the one proposed for vision (Rauschecker, 1998). Areas at the junction of superior temporal and inferior parietal cortices may be well-situated anatomically for binding the two streams. Animal models suggest that supra-sylvian and infra-sylvian

regions are massively interconnected (Petrides & Pandya, 2009). It is possible that the spatial processing in the dorsal stream is sufficiently abstract so as to play a role in different domains. Luria (1970) described patients with lesions near the junction(s) of temporal, parietal, and occipital lobes who showed a variety of deficits, including the inability to distinguish between phrases like "brother's father" and "father's brother" and poor comprehension of complex sentences, mathematical calculation, and spatial orientation. More recently, BA 39 has been implicated as a possible common substrate for arithmetic, relational reasoning, and sentence comprehension (Baldo, Bunge, Wilson, & Dronkers, 2010; Baldo & Dronkers, 2007). It has also been proposed that a nonlinguistic spatial representation might underlie our understanding of thematic relations in language (Coslett, 1999; Chatterjee, Maher, Rothi, & Heilman, 1995). These speculations can be tested by future studies evaluating whether damage to the junction of temporal and parietal areas leads to broader spatial and/or relational deficits, as measured by nonlinguistic tasks and the comprehension of other relational linguistic structures.

VLPFC

Earlier studies reported evidence from Broca's aphasia in favor of a critical role for Broca's area in the syntactic aspects of comprehension (Caramazza & Zurif, 1976). Such studies seldom had the anatomical precision required to make brain–function correlations. Several linking assumptions in inferring this brain–function relation have been called into question. First, not all Broca's aphasics have damage to Broca's area (Dronkers, 2000). Second, not all Broca's aphasics have impaired syntactic comprehension (Berndt, Mitchum, & Haendiges, 1996). Last but not the least, patients with Broca's area damage do not invariably show syntactic comprehension deficits. One recent study tested comprehension of passives in 38 agrammatic aphasics with verified damage to Broca's area (Caramazza, Capasso, Capitani, & Miceli, 2005). More patients performed better than chance than would be expected if Broca's area were critical for the comprehension of noncanonical sentences. Collectively, these findings call into question the proposed link from Broca's aphasia to Broca's area damage to poor syntactic comprehension. A look at the Broca's aphasics in our sample offers a similar cautionary note. Some of the patients with extensive Broca's area (BA 44) damage performed well above chance in our comprehension task. Temporo-parietal (BA 39) damage better separated the poor comprehenders from the good ones.

Neuroimaging studies are equivocal about the correct interpretation of VLPFC activation during comprehension tasks. Although some authors propose that VLPFC, particularly Brocas's area, supports critical syntactic operations (Ben-Shachar et al., 2003), others have attributed a less central role to Broca's area during sentence comprehen-

sion. For example, Caplan, Stanczak, and Waters (2008) suggest that the initial assignment of thematic roles based on sentence structure is carried out by left temporal areas and that left frontal areas may be involved only in the later checking of those roles under certain conditions. Fiebach et al. (2005, p. 89) state: "We suggest that BA 44 is recruited mainly in cases when syntactic information has to be maintained temporarily in working memory. Parsing processes that are more computational in nature and temporally more circumscribed might be carried out partly in other brain regions."

Our interpretation of the negative VLPFC results from the current study is consistent with such interpretations. We suggest that Broca's area does not play a task-independent, core syntactic parsing role in comprehension. Instead, it may support the cognitive control and/or working memory resources that are often but not always associated with sentence comprehension. This hypothesis is consistent with the fact that, in the current study, the only suggestion of a preferential VLPFC effect was found when we contrasted two-proposition with one-proposition sentences— a manipulation that has been hypothesized to tap working memory and other task-related resources (Waters et al., 1998).

There is widespread evidence that VLPFC is involved in the selection of a task-relevant representation from among mutually incompatible alternatives (Petrides, 2005; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997). In sentence comprehension, VLPFC involvement may be particularly important, even crucial, in cases where conflicting cues (e.g., semantics and syntax) must be resolved to converge on the correct interpretation. Future research can determine whether patients with damage in Broca's area will be particularly impaired in such tasks. For the present, results from our VLSM and regional analyses suggest that VLPFC does not play a major role in successful performance in a widely used reversible sentence comprehension test.

Concluding Remarks

Our analyses show a robust association between impaired reversible sentence comprehension and damage in left temporo-parietal areas. In lesion studies, areas may show up as important either because they play a direct causal role in the behavior of interest or because they are functionally affected by disrupted connectivity from elsewhere. Our VLSM and regional analyses are congruent in suggesting that a functional left temporo-parietal cortex is necessary for normal reversible sentence comprehension. Nevertheless, both are "static" techniques that allow us to correlate behavioral impairments with brain damage. A fuller picture of how sentence comprehension unfolds in the brain requires the integration of several methods, including electrophysiological techniques that have higher temporal resolution and functional connectivity studies that shed light on the transfer of information between different areas.

In focusing on the temporo-parietal region, we do not mean to suggest that it is the sole area responsible for sentence comprehension under all circumstances. As mentioned above, some stimuli or tasks may tap working memory and cognitive control resources more than those used here. In such situations, VLPFC-supported executive functions may play a critical role. Our comprehension task also did not test fine-grained lexical processing. A small set of words—eight unique nouns and four unique verbs—was repeated again and again in different sentences. A functioning anterior temporal lobe may be crucial when fine-grained lexico-semantic distinctions or the integration of meaning from multiple words is important for good performance (Warren et al., 2009). It is a virtual certainty that a complex task such as sentence comprehension involves a distributed network of regions subserving multiple functions (Caplan et al., 2007; Dronkers et al., 2004). Our results suggest that the temporo-parietal region may be critical for a particular aspect of sentence comprehension, namely assigning thematic roles using sentence structure.

Acknowledgments

Notes

1. Actives with prepositional phrases had the same syntax to thematic role mappings as actives but were more comparable in length to the relative clause sentence types. Throughout the manuscript, we use "actives" to refer only to bare actives (without prepositional phrases).
2. Standard STM span measures may tap both semantic and phonological STM. Because temporo-parietal areas have been specifically tied to phonological STM, we used a task that has been used to measure this component (Freedman & Martin, 2001).
3. Because VLSM compares two groups at each voxel (lesioned and nonlesioned), imposing such a threshold helps to ensure sufficient numbers in each group for obtaining a stable estimate of variance in the behavioral scores. The choice of threshold is arbitrary.
4. We obtained similar results when we included actives with prepositional phrases in the canonical sentence group (mean = 85.7%, $F(1, 78) = 97.1$, $p < .001$). We did not include locatives in these analyses because there is no clear consensus on the thematic roles and their canonical mappings to syntax for such structures.
5. Separate VLSM analyses for canonical and noncanonical scores also support this conclusion. Suprathreshold peak voxels were found in BA 39 after covarying out rhyme probe span (canonical: peak MNI, -38, -53, 31; noncanonical: peak MNI, -55, -54, 31) and NWRep scores (canonical: peak MNI, -38, -53, 31; noncanonical: peak MNI, -54, -55, 35 ($t = 5.19$).

REFERENCES

Amici, S., Brambati, S. M., Wilkins, D. P., Ogar, J., Dronkers, N. L., Miller, B. L., et al. (2007). Anatomical correlates of sentence comprehension and verbal working memory in neurodegenerative disease. *Journal of Neuroscience, 27,* 6282.

Avants, B. B., Schoenemann, P. T., & Gee, J. C. (2006). Lagrangian frame diffeomorphic image registration: Morphometric comparison of human and chimpanzee cortex. *Medical Image Analysis, 10,* 397–412.

Baldo, J. V., Bunge, S. A., Wilson, S. M., & Dronkers, N. F. (2010). Is relational reasoning dependent on language? A voxel-based lesion symptom mapping study. *Brain and Language, 113,* 59–64.

Baldo, J. V., & Dronkers, N. F. (2007). Neural correlates of arithmetic and language comprehension: A common substrate? *Neuropsychologia, 45,* 229–235.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., et al. (2003). Voxel-based lesion symptom mapping. *Nature Neuroscience, 6,* 448–450.

Ben-Shachar, M., Hendler, T., Kahn, I., Ben-Bashat, D., & Grodzinsky, Y. (2003). The neural reality of syntactic transformations: Evidence from functional magnetic resonance imaging. *Psychological Science, 14,* 433–440.

Ben-Shachar, M., Palti, D., & Grodzinsky, Y. (2004). Neural correlates of syntactic movement: Converging evidence from two fMRI experiments. *Neuroimage, 21,* 1320–1336.

Berndt, R. S., Mitchum, C. C., & Haendiges, A. N. (1996). Comprehension of reversible sentences in "agrammatism": A meta-analysis. *Cognition, 58,* 289–308.

Caplan, D., Alpert, N., & Waters, G. (1998). Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience, 10,* 541–552.

Caplan, D., Alpert, N., & Waters, G. (1999). PET studies of syntactic processing with auditory sentence presentation. *Neuroimage, 9,* 343–351.

Caplan, D., Stanczak, L., & Waters, G. (2008). Syntactic and thematic constraint effects on blood oxygen level dependent signal correlates of comprehension of relative clauses. *Journal of Cognitive Neuroscience, 20,* 643–656.

Caplan, D., Vijayan, S., Kuperberg, G., West, C., Waters, G., Greve, D., et al. (2001). Vascular responses to syntactic processing: Event-related fMRI study of relative clauses. *Human Brain Mapping, 15,* 26–38.

Caplan, D., Waters, G., Kennedy, D., Alpert, N., Makris, N., DeDe, G., et al. (2007). A study of syntactic processing in aphasia II: Neurological aspects. *Brain and Language, 101,* 151–177.

Caramazza, A., Capasso, R., Capitani, E., & Miceli, G. (2005). Patterns of comprehension performance in agrammatic Broca's aphasia: A test of the Trace Deletion Hypothesis. *Brain and Language, 94,* 43–53.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language, 3,* 572–582.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113,* 234–272.

Chatterjee, A., Maher, L. M., Rothi, L. J. G., & Heilman, K. M. (1995). Syntactic thematic role assignment: The use of a temporal-spatial strategy. *Brain and Language, 49,* 125–139.

Coslett, H. B. (1999). Spatial influences on motor and language function. *Neuropsychologia, 37,* 695–706.

Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. J. S. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain, 126,* 1193–1201.

Curtiss, S., & Yamada, J. (1988). The Curtiss–Yamada comprehensive language evaluation (CYCLE). Unpublished test. University of California, Los Angeles, CA.

den Ouden, D. B., Fix, S., Parrish, T. B., & Thompson, C. K. (2009). Argument structure effects in action verb naming in static and dynamic conditions. *Journal of Neurolinguistics, 22,* 196–215.

Dronkers, N. F. (2000). The gratuitous relationship between Broca's aphasia and Broca's area. *Behavioral and Brain Sciences, 23,* 30–31.

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition, 92,* 145–177.

Fiebach, C. J., Schlesewsky, M., Lohmann, G., von Cramon, D. Y., & Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: Syntactic integration vs. syntactic working memory. *Human Brain Mapping, 24,* 79–91.

Freedman, M. L., & Martin, R. C. (2001). Dissociable components of short-term memory and their relation to long-term learning. *Cognitive Neuropsychology, 18,* 193–226.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage, 15,* 870–878.

Grodzinsky, Y., & Santi, A. (2008). The battle for Broca's region. *Trends in Cognitive Sciences, 12,* 474–480.

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

Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., & Evans, A. C. (1998). Enhancement of MR images using registration for signal averaging. *Journal of Computer Assisted Tomography, 22,* 324.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science, 274,* 114–116.

Kertesz, A. (1982). *Western aphasia battery test manual.* San Antonio, TX: Psychological Corp.

Kimberg, D. Y., Coslett, H. B., & Schwartz, M. F. (2007). Power in voxel-based lesion symptom mapping. *Journal of Cognitive Neuroscience, 19,* 1067–1080.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience, 9,* 920–933.

Luria, A. R. (1970). *Traumatic aphasia.* The Hague: Mouton & Co.

Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex, 13,* 1362–1368.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 5,* 263–281.

Petrides, M. (2005). Lateral prefrontal cortex: Architectonic and functional organization. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 360,* 781–795.

Petrides, M., & Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biology, 7,* e1000170.

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology, 8,* 516–521.

Rorden, C., & Karnath, H. O. (2004). Using human brain lesions to infer function: A relic from a past era in the fMRI age? *Nature Reviews Neuroscience, 5,* 813–819.

Saffran, E. M., Schwartz, M. F., Linebarger, M. C., Martin, N., & Bochetto, P. (1988). Philadelphia comprehension battery. Unpublished test battery.

Schwartz, M. F., Brecher, A. R., Whyte, J., & Klein, M. G. (2005). A patient registry for cognitive rehabilitation research: A strategy for balancing patients' privacy rights with researchers' need for access. *Archives of Physical Medicine and Rehabilitation, 86,* 1807–1814.

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Faseyitan, O., Brecher, A., Dell, G. S., et al. (2009). Anterior temporal involvement in semantic word retrieval: Voxel-based lesion symptom mapping evidence from aphasia. *Brain, 132,* 3411–3427.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123,* 2400–2406.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245–251.

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language, 52,* 452–473.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, U.S.A., 94,* 14792–14797.

Warren, J. E., Crinion, J. T., Ralph, M. A. L., & Wise, R. J. S. (2009). Anterior temporal lobe connectivity correlates with functional outcome after aphasic stroke. *Brain, 132,* 3428–3442.

Wartenburger, I., Heekeren, H. R., Burchert, F., Heinemann, S., De Bleser, R., & Villringer, A. (2004). Neural correlates of syntactic transformations. *Human Brain Mapping, 22,* 72–81.

Waters, G. S., Rochon, E., & Caplan, D. (1998). Task demands and sentence comprehension in patients with dementia of the Alzheimer's type. *Brain and Language, 62,* 361–397.

Wernicke, C. (1874). *Der aphasische symptomencomplex.* Breslau: Cohn and Weigert.

Wise, R. J. S. (2003). Language systems in normal and aphasic human subjects: Functional imaging studies and inferences from animal studies. *British Medical Bulletin, 65,* 95–119.

Yokoyama, S., Watanabe, J., Iwata, K., Ikuta, N., Haji, T., Usui, N., et al. (2007). Is Broca's area involved in the processing of passive sentences? An event-related fMRI study. *Neuropsychologia, 45,* 989–996.