

Functional specificity for high-level linguistic processing in the human brain

Evelina Fedorenko^{a,1}, Michael K. Behr^a, and Nancy Kanwisher^{a,b,1}

^aBrain and Cognitive Sciences Department and ^bMcGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Nancy Kanwisher, August 9, 2011 (sent for review July 13, 2011)

Neuroscientists have debated for centuries whether some regions of the human brain are selectively engaged in specific high-level mental functions or whether, instead, cognition is implemented in multifunctional brain regions. For the critical case of language, conflicting answers arise from the neuropsychological literature, which features striking dissociations between deficits in linguistic and nonlinguistic abilities, vs. the neuroimaging literature, which has argued for overlap between activations for linguistic and nonlinguistic processes, including arithmetic, domain general abilities like cognitive control, and music. Here, we use functional MRI to define classic language regions functionally in each subject individually and then examine the response of these regions to the nonlinguistic functions most commonly argued to engage these regions: arithmetic, working memory, cognitive control, and music. We find little or no response in language regions to these nonlinguistic functions. These data support a clear distinction between language and other cognitive processes, resolving the prior conflict between the neuropsychological and neuroimaging literatures.

functional specialization | modularity | functional MRI | sentence understanding

Language is our signature cognitive skill, uniquely and universally human. Does language rely on dedicated cognitive and neural machinery specialized for language only, or is it supported by the same circuits that enable us to perform arithmetic, hold information in working memory, or appreciate music?

The neuropsychological literature features striking dissociations between deficits in linguistic and nonlinguistic abilities (1–6). In contrast, prior neuroimaging work has suggested that brain regions that support language—especially regions on the lateral surface of the left frontal lobe—also support nonlinguistic functions (7–23). However, the neuroimaging claims are based largely on observations of activation for a nonlanguage task in or near a general brain region previously implicated in language (e.g., Broca's area), without a direct demonstration of overlap of linguistic and nonlinguistic activations in the same subjects. Such findings do not suffice to demonstrate common brain regions for linguistic and nonlinguistic functions in the brain. Even when two tasks are compared in the same group of subjects, standard functional MRI group analysis methods can be deceptive: two different mental functions that activate neighboring but non-overlapping cortical regions in every subject individually can produce overlapping activations in a group analysis, because the precise locations of these regions vary across subjects (24–26), smearing the group activations. Definitively addressing the question of neural overlap between linguistic and nonlinguistic functions requires examining overlap within individual subjects (27, 28), a data analysis strategy that has almost never been applied in neuroimaging investigations of high-level linguistic processing.

Here, we used functional MRI to test the functional specificity of brain regions engaged in language processing by identifying brain regions that are sensitive to high-level linguistic processing in each subject individually and then asking whether these same regions are also engaged when the subject performs nonlinguistic

tasks. Specifically, we examined the response of high-level language regions to seven nonlinguistic tasks that tap the cognitive functions most commonly argued to recruit the language network: exact arithmetic (experiment 1), working memory (WM; experiments 2 and 3), cognitive control (experiments 4–6), and music (experiment 7).

For the language localizer task, participants read sentences and strings of pronounceable nonwords and decided whether a subsequent probe (a word or nonword, respectively) appeared in the preceding stimulus (Fig. 1). The sentences > nonwords contrast targets brain regions engaged in high-level linguistic processing (sentence understanding), including lexical and combinatorial processes but excluding phonological and other lower-level processes.* This localizer was previously shown (29) to robustly and reliably identify each of the key brain regions previously implicated in high-level linguistic processing (30, 31) (Fig. 2 and Fig. S1) in at least 80% of subjects individually and to be robust to changes in stimulus modality (visual/auditory) (see also ref. 32), specific stimuli, and task (passive reading vs. reading with a probe task, which is used here). Furthermore, unlike the language contrasts used in some prior studies (17, 21), this contrast does not confound linguistic processing and general cognitive effort, because the memory probe task is more difficult in the control (nonwords) condition.

For all nonlinguistic functions, we chose tasks and contrasts commonly used in previous published studies (Fig. 1 and *Materials and Methods*). Experiments 1–6 each contrasted a condition that was more demanding of the nonlinguistic mental process in question and a closely matched control condition that was less demanding of that mental process (see Table S3 for behavioral results). Engagement of a given region in a given mental process would be shown by a higher response in the difficult condition than the easy condition. Given previous findings (33–35), we expected these tasks to activate frontal and parietal structures bilaterally. Experiment 7 contrasted listening to intact vs. scrambled music. Engagement of a given region in music would be shown by a higher response to intact than scrambled music. Given previous findings (3, 20, 36), we expected this contrast to activate frontal and temporal brain regions. The critical question is whether the activations for nonlinguistic task contrasts will overlap with activations for the language task.

Author contributions: E.F. and N.K. designed research; E.F. and M.K.B. performed research; E.F. analyzed data; and E.F. and N.K. wrote the paper.

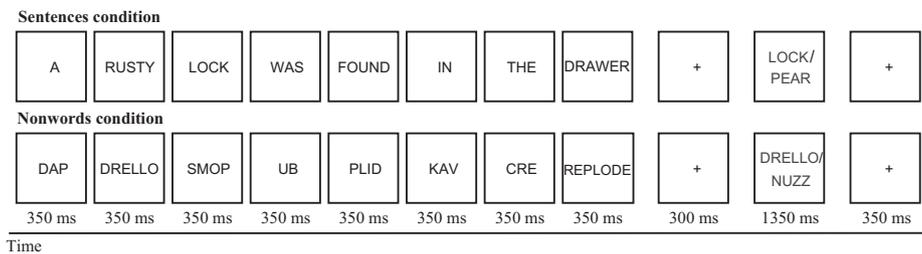
The authors declare no conflict of interest.

¹To whom correspondence may be addressed. E-mail: evelina9@mit.edu or ngk@mit.edu.

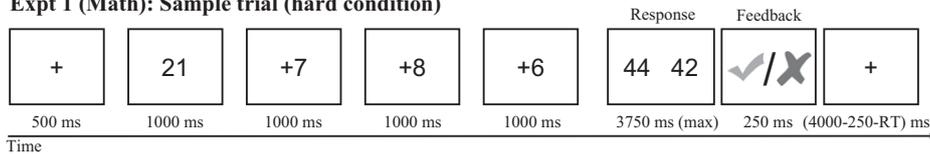
This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1112937108/-DCSupplemental.

*Of course, the question of the relationship between linguistic and nonlinguistic processes extends to other aspects of language (e.g., sound-level processes), which we are not targeting with the current language localizer task. However, most claims about the overlap between linguistic and nonlinguistic processes have concerned (i) syntactic processing, which is included in our functional contrast, and (ii) brain regions, which are robustly activated by our localizer (e.g., Broca's area).

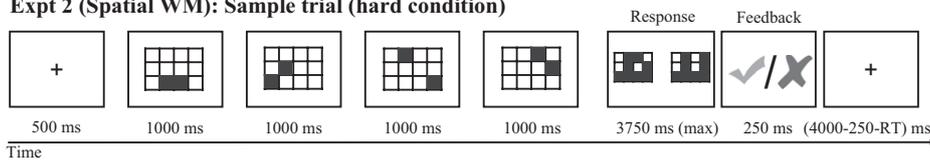
Language localizer: Sample trials



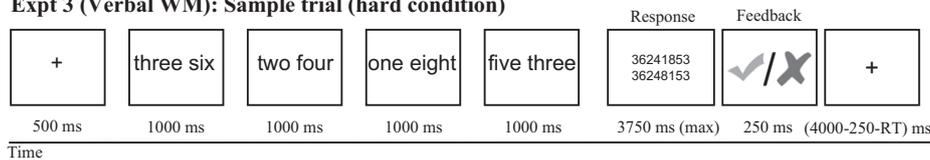
Expt 1 (Math): Sample trial (hard condition)



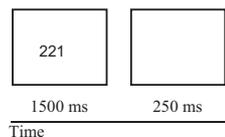
Expt 2 (Spatial WM): Sample trial (hard condition)



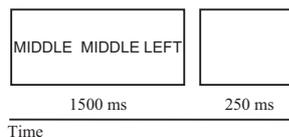
Expt 3 (Verbal WM): Sample trial (hard condition)



Expt 4 (MSIT): Sample trial (hard condition)



Expt 5 (vMSIT): Sample trial (hard condition)



Expt 6 (Stroop): Sample trial (hard condition)

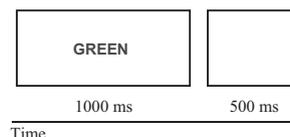


Fig. 1. Sample trials for the localizer task and nonlinguistic tasks in experiments 1–6. In the math task, participants added smaller vs. larger addends. In the spatial/verbal WM tasks, participants remembered four vs. eight locations/digit names. In the multisource interference task (MSIT), participants saw triplets of digits and had to press a button corresponding to the identity of the nonrepeated digit. The hard condition involves two kinds of conflict: between the spatial position of the target response and the position of the response button and between the target response and the distracter elements. The verbal MSIT (vMSIT) task was a slight variant of the MSIT, where words were used instead of digits. In the Stroop task, participants overtly named the font color vs. noncolor adjectives. In the music task, participants listened to unfamiliar Western tonal pieces vs. the scrambled versions of these pieces (*Materials and Methods* has details).

Results

Language Activations. As expected, reading sentences elicited a reliably stronger response than reading nonwords in each region of interest (ROI) (*Materials and Methods*, Fig. 2, black and gray bars, and Fig. S1, black and gray bars) (here and throughout the study, responses were estimated using data from left-out runs not used for defining the ROIs).

Overlap Between Language and Nonlinguistic Tasks: ROI Analyses.

Fig. 2 and Tables S1 and S2 show our key results: the response in each of the left hemisphere language-sensitive ROIs during each of the seven nonlinguistic tasks (Fig. S1 has the responses of right hemisphere and cerebellar regions). Although each nonlinguistic task produced robust activations at the individual subject level in nearby brain regions (Fig. 3), they had little effect in the language ROIs. The music contrast did produce some response in several language ROIs (*SI Text*), suggesting possibly overlapping circuits; however, these effects were weak overall, and none survived cor-

rection for multiple comparisons. One frontal language ROI left middle frontal gyrus (LMFG) showed a significant response to verbal working memory after correcting for the number of ROIs; this region is, thus, not specialized for sentence understanding. However, no other ROIs showed reliable responses to any of the nonlinguistic tasks, indicating a striking degree of functional specificity for language processing.

In several ROIs, some of the nonlinguistic tasks resulted in deactivation relative to the fixation baseline. These patterns simply indicate that the region in question was more active during rest (when participants are plausibly engaged in daydreaming, presumably sometimes including verbal thinking) than during the task in question.

Overlap Between Language and Nonlinguistic Tasks: Whole-Brain Analyses. ROI-based analyses afford great statistical power and reduce the blurring of functional activations inevitable in group analyses. However, ROI analyses can obscure functional heterogeneity within an ROI or miss important overlapping regions

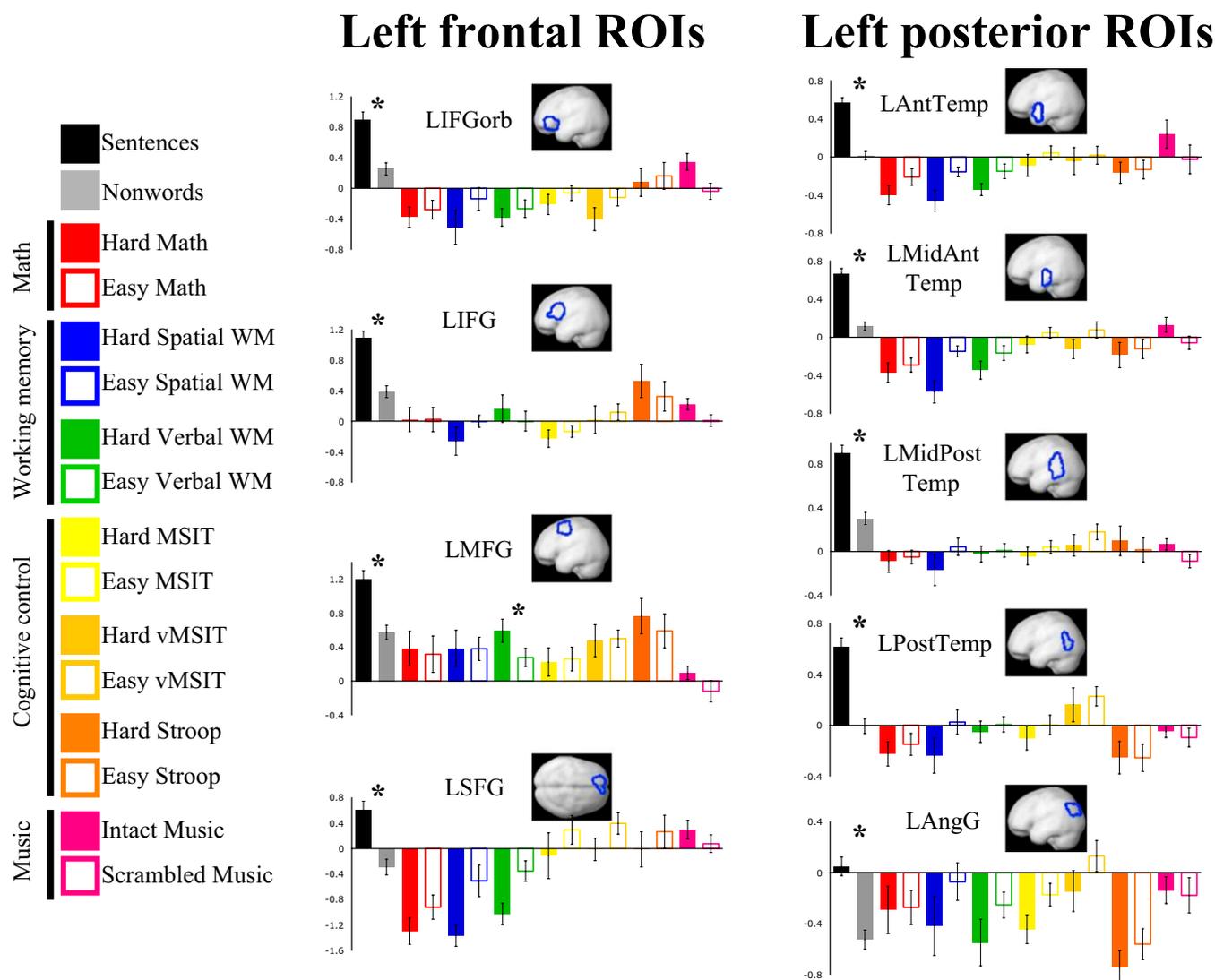


Fig. 2. Responses of the left hemisphere language ROIs to the language task (black, sentences; gray, nonwords; estimated using data from left-out runs) and the nonlinguistic tasks [shown in different colors, with the solid bar showing the response to the hard (experiments 1–6)/intact music (experiment 7) condition and the nonfilled bar showing the response to the easy/scrambled music condition]. Individual subjects' ROIs were defined by intersecting the functional parcels (shown in blue contours) with each subject's thresholded ($P < 0.001$, uncorrected) activation map for the language localizer. The functional parcels were generated a priori based on a group-level representation of localizer activations in an independent set of subjects (29). Asterisks indicate significant effects (Bonferroni-corrected for the number of regions). The sentences > nonwords effect was highly significant ($P < 0.0001$) in each functional ROI. The only significant effect for the nonlinguistic tasks was the hard > easy verbal WM effect in the LMFG functional ROI ($P < 0.005$). Tables S1 and S2 have detailed statistics.

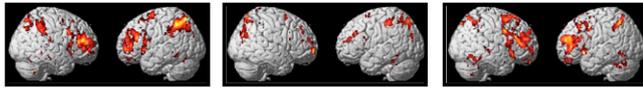
outside the borders of the ROI. We, therefore, conducted additional analysis that did not use our language ROIs but instead, searched the whole brain for any regional consistency across subjects in the overlap between language activations and each of the nonlinguistic tasks (*Materials and Methods*).

At the threshold used for the ROI analyses, no spatially systematic regions were detected in which language activations overlapped with any of the seven nonlinguistic tasks, confirming the overall picture of language specificity found in the ROI analyses. [This threshold is not too stringent to discover true overlap when it exists; at the same threshold, we discovered robust overlap regions in the frontal and parietal lobes between Stroop and the multisource interference task (MSIT) cognitive control task in a subset of the subjects run on both tasks.]

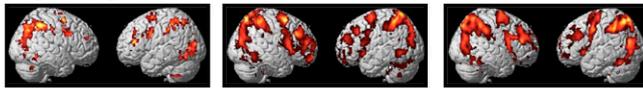
To maximize our chances of discovering activation overlap, we performed the same analysis again but now with very liberally thresholded individual maps. This analysis discovered three overlap regions. Two of these regions were fully consistent with

the results of the ROI analyses: one region was located within the LMFG language ROI and responded to both language and verbal WM, and another region was within the LMidPostTemp language ROI and responded to language and music (*SI Text* and Fig. S2). The third region, present in 85% of the subjects, emerged in the posterior part of our left inferior frontal gyrus (LIFG) language ROI and responded to both language and verbal WM (a similar region emerged in the analysis of the overlap of language and Stroop activations, although it was present in only 71% of the subjects) (Fig. 4). The existence of this overlap region in posterior LIFG is consistent with the weak trends that we found in the LIFG language ROI for verbal WM and Stroop. Furthermore, it is consistent with evidence for fine-grained cytoarchitectonic subdivisions within Broca's area (37, 38). Nevertheless, a complementary analysis shows that a portion of the LIFG language ROI is selective for high-level linguistic processing and not engaged in either verbal WM or Stroop (Fig. 4).

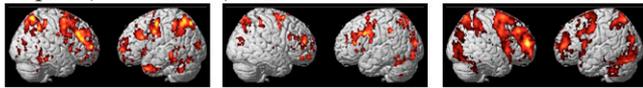
Expt 1 (Math)



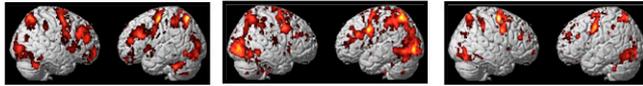
Expt 2 (Spatial WM)



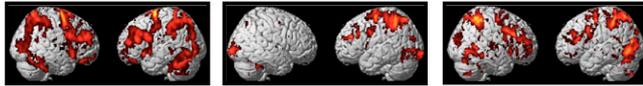
Expt 3 (Verbal WM)



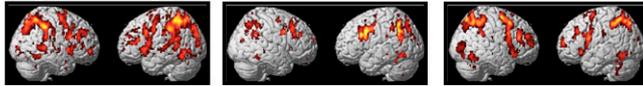
Expt 4 (MSIT)



Expt 5 (vMSIT)



Expt 6 (Stroop)



Expt 7 (Music)

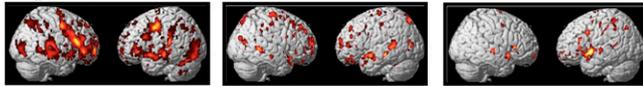


Fig. 3. Sample individual activation maps (thresholded at $P < 0.001$, uncorrected) for the nonlinguistic tasks showing three representative subjects from each experiment. For experiments 1–6, the contrast is hard > easy; for experiment 7, the contrast is intact > scrambled.

Discussion

The current results show that most of the key cortical regions engaged in high-level linguistic processing are not engaged by mental arithmetic, general working memory, cognitive control, or musical processing, indicating a high degree of functional specificity in the brain regions that support language. The one exception is a brain region located in the left middle frontal gyrus that seems to support both high-level linguistic processing and verbal working memory.

The ROI corresponding roughly to Broca's area (LIFG) (Fig. 2) has been a major target of prior research, and we chose nonlinguistic tasks that could test the main hypotheses that have been proposed about the possible nonlinguistic function(s) of this region. Our data are difficult to reconcile with the widely favored hypothesis that this region is engaged in domain general cognitive control (8, 17); although we do see a weak response to the Stroop task in this region, we find no hint of activation for either of the other two cognitive control tasks [the MSIT task (39) and its variant, verbal MSIT (vMSIT)]. Furthermore, although this region does respond more strongly to intact than scrambled musical stimuli (consistent with some previous claims) (21), the response to intact music is barely above the fixation baseline (and lower than the response to the language localizer control task), making it difficult to argue that this region supports general structural processing of complex stimuli that unfold over time (9, 20, 21).

The nature of the computations in the posterior part of the LIFG language ROI that seems to support not only sentence understanding but also verbal WM and the mental processes re-

quired for the Stroop task (Fig. 4) remains an open question. This region—like the language ROI encompassing it (LIFG ROI)—is unlikely to support domain general cognitive control, because no overlap regions emerged in posterior LIFG between language and the other two cognitive control tasks (MSIT and vMSIT), consistent with the lack of a response to these tasks in the LIFG language ROI. Moreover, the fact that this region responds to the verbal WM task, which is not traditionally considered to tax cognitive control processes, suggests that this region may instead support demanding cognitive tasks when they involve storing or manipulating verbal representations. It is also possible, however, that, at higher spatial resolution, we would discover that this apparent overlap is artifactual, merely reflecting the proximity of the language-sensitive cortical regions to those regions that support demanding cognitive tasks (ref. 40 has a similar demonstration in the ventral visual cortex).

More generally, the current data pose a challenge for hypotheses that brain regions that support high-level aspects of language also (i) represent and manipulate exact quantity information (7, 22, 23), (ii) contribute to language processing through domain general working memory or cognitive control mechanisms (8, 17), or (iii) process structured sequences unfolding over time regardless of their content (9, 16, 20).

Note that the brain regions investigated here are surely not the only ones engaged in or necessary for language processing. Understanding language requires perceiving linguistic input either auditorily (hence, requiring auditory cortex) or visually (hence, requiring visual cortex), and producing language requires parts of motor cortex. In addition, frontal and parietal brain regions engaged in a very domain general fashion in many demanding cognitive tasks (33–35) may be necessary for many aspects of linguistic function. Thus, language processing probably requires both the very specialized regions studied here and more domain general regions.

Of course, discovering that a brain region is selectively engaged in linguistic processing is just a first step. Ultimately, we want to know the precise computations conducted and the representations extracted in each of these regions as well as how different components of the language system work together to enable us to produce and understand language. Existing frameworks (41–44) provide an important foundation, but future work should strive to develop computationally precise and testable hypotheses for each of the core language regions.

To conclude, discoveries of linguistic specificity provide answers to the longstanding question of whether high-level cognitive functions are supported by dedicated neural machinery (45) and provide important clues into the mental operations supported by these regions. Future work will (i) test the engagement of regions that support high-level linguistic processing in yet other nonlinguistic processes (e.g., amodal conceptual processing, action representation, or hierarchical sequence processing) as well as examine other language regions that support phonological and discourse-level processes, (ii) identify the precise linguistic computations conducted in each of these regions, and (iii) characterize the interactions between these regions and other parts of the brain that enable uniquely human cognition.

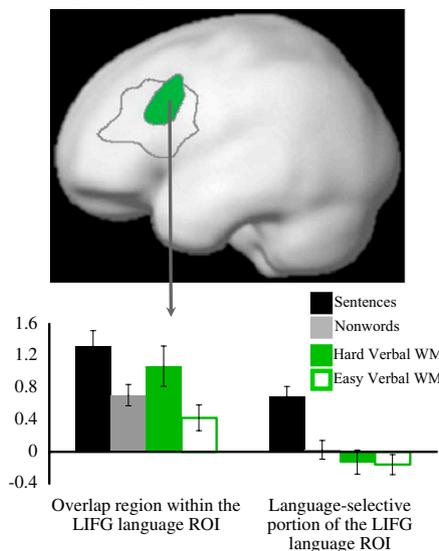
Materials and Methods

Participants. Forty-eight right-handed participants (31 females) from Massachusetts Institute of Technology (MIT) and the surrounding community were paid for their participation. All were native speakers of English between the ages of 18 and 40 y, had normal or corrected to normal vision, and were naive as to the purposes of the study. All participants gave informed consent in accordance with the requirements of the Internal Review Board at MIT.

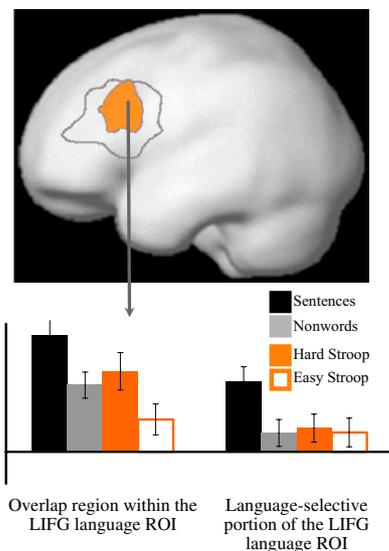
Design. Each participant was run on the language localizer task (which included the sentences condition and the pronounceable nonwords condition) and one or more nonlinguistic tasks. Each nonlinguistic task included two conditions (hard and easy for tasks in experiments 1–6 and intact music and scrambled music in experiment 7) (Fig. 1).

Fig. 4. Key results from the whole-brain overlap analysis with very liberally thresholded ($P < 0.05$, uncorrected) individual activation maps (*Materials and Methods* has details). Similar overlap regions—in the posterior part of the LIFG language ROI (whose outline is shown in gray above)—emerged in the analyses of language and verbal WM (green; present in 85% of the subjects with an average size of 76 voxels) and language and Stroop (orange; present in 71% of the subjects with an average size of 95 voxels). The left set of bars in each panel shows the responses of these overlap regions to the four conditions (estimated in left-out runs). The right set of bars in each panel shows the results of a complementary analysis where we examined voxels within the LIFG language ROI that respond to the language localizer contrast (at $P < 0.001$, uncorrected) but not to the non-linguistic task contrast (at $P < 0.1$). Responses to the four conditions were again estimated using left-out runs. In each dataset, a substantial proportion of voxels (127/355 voxels, on average, in the language-Verbal Working Memory dataset and 163/410 voxels in the language-Stroop dataset) showed replicable selectivity for the language task, suggesting that a portion of the LIFG language ROI is selectively engaged in sentence understanding.

Language and Verbal WM



Language and Stroop



Language localizer. Participants read sentences and pronounceable nonword lists and presented one word/nonword at a time. After each stimulus, a memory probe appeared, and participants decided whether the probe word/nonword appeared in the preceding stimulus.

The localizer is available to download at http://web.mit.edu/evelina9/www/funcloc/funcloc_localizers.html.

Nonlinguistic tasks. In experiment 1 (math), participants ($n = 11$) saw a number (11–30) and added three addends to it (of sizes two to four or six to eight in the easy and hard conditions, respectively).¹ After each trial, participants had to choose the correct sum in a two-choice forced choice question (the incorrect sum deviated by one to two). Participants were told whether they answered correctly. Each run (consisting of 10 34-s-long experimental and 6 16-s-long fixation blocks) lasted 436 s. Each participant did two to three runs. (The same timing was used in experiments 2 and 3.) In experiment 2 (spatial WM), participants ($n = 12$) saw a 3×4 grid and kept track of four or eight locations in the easy and hard conditions, respectively. After each trial, participants had to choose the grid with the correct locations in a two-choice question (the incorrect grid contained one to two wrong locations). In experiment 3 (verbal WM), participants ($n = 13$) kept track of four- or eight-digit names in the easy and hard conditions, respectively. Digits were presented as words (e.g., three) to prevent chunking (e.g., to prevent 3 5 from being encoded as thirty-five). After each trial, participants had to choose the correct digit sequence in a two-choice question (in the incorrect sequence, the identity of one or two digits, with respect to serial position, was incorrect). In experiment 4 (MSIT), participants ($n = 15$) saw triplets of digits (possible digits included zero, one, two, and three) and pressed a button (one, two or three; laid out horizontally on a button box) corresponding to the identity of the nonrepeated digit. In the easy condition, the position of the nonrepeated digit corresponded to the position of the response button, and the other digits were not possible responses (e.g., 100); in the hard condition, the position of the nonrepeated digit did not correspond to the position of the button, and the other digits were possible responses (e.g., 212). The timing was as described in ref. 39. Each participant did three to five runs. Experiment 5 (vMSIT, $n = 12$) was identical to experiment 4, except that digits (zero, one, two, and three) were replaced with words (none, left, middle, and right) to see whether more verbal representations would lead to greater overlap with the language activations. In experiment 6 (Stroop), participants ($n = 14$) saw a word and overtly named the color of the word's font. In the easy condition, the words were noncolor adjectives (huge, close, and guilty) matched to the color adjectives in length and lexical frequency; in the hard condition, the words were color adjectives (blue, green, and

yellow), and in one-half of the trials in each block, the font color did not match the color that the word indicated. Experimental and fixation blocks lasted 18 s each. Each run (16 experimental and 5 fixation blocks) lasted 378 s. Each participant did two to three runs. In experiment 7 (music), participants ($n = 12$) heard 64 unfamiliar musical pieces (pop/rock music from the 1950s and 1960s) over the scanner-safe headphones. To scramble musical structure, we altered the pitch and timing of musical notes in Musical Instrument Digital Interface representations of the pieces. For each piece, a random number of semitones between -3 and 3 was added to the pitch of each note to make the pitch distribution approximately uniform. The resulting pitch values were randomly reassigned to the notes of the piece to remove the contour structure. To remove the rhythmic structure, the note onsets were jittered by up to 1 beat duration (uniformly distributed), and the note durations were randomly reassigned. The resulting piece had component sounds similar to the intact music but lacked musical structure. (The experiment included two additional conditions, with only the pitch contour or rhythm scrambled, but we focus on the intact and scrambled conditions here.) For eight participants, the task was passive listening (to help participants stay awake, we asked them to press a button after each piece); the last four participants were asked the question, "How much do you like this piece?" after each stimulus. Because the activation patterns were similar across the two tasks, we collapsed the data from these two groups of participants for the current analyses. Experimental and fixation blocks lasted 24 and 16 s, respectively. Each run (16 experimental and 5 fixation blocks) lasted 464 s. Each participant did four to five runs.

Data Acquisition. Structural and functional data were collected on the whole-body 3-Tesla Siemens Trio scanner at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. T1-weighted structural images were collected in 128 axial slices with 1.33-mm isotropic voxels [repetition time (TR) = 2,000 ms and echo time (TE) = 3.39 ms]. Functional, blood oxygenation level-dependent data were acquired in $3.1 \times 3.1 \times 4$ -mm voxels (TR = 2,000 ms and TE = 30 ms) in 32 near-axial slices. The first 4 s of each run were excluded to allow for steady state magnetization. The scanning session included several functional runs. The order of conditions was counterbalanced; each condition was equally likely to appear in the earlier vs. later parts of each run and was as likely to follow every other condition as it was to precede it.

Statistical Analysis. MRI data were analyzed using SPM5 (<http://www.fil.ion.ucl.ac.uk/spm>) and custom matlab scripts (available from <http://web.mit.edu/evelina9/www/funcloc.html>). Each subject's data were motion-corrected and then, they were normalized onto a common brain space (the Montreal Neurological Institute template) and resampled into 2-mm isotropic voxels. Data were smoothed using a 4-mm Gaussian filter and high pass-filtered (at 200 s). Language-sensitive ROIs were defined by intersecting (\cap) the parcels generated

¹Overlapping sets of subjects were run on different nonlinguistic tasks; hence, the total number of participants ($n = 48$) is less than the sum of all of the numbers of participants across the seven experiments.

based on a group-level representation of the localizer data in a separate experiment (the parcels' outlines are shown in blue in Fig. 2 and Fig. S1) and (ii) individual subjects' localizer maps thresholded at $P < 0.001$, uncorrected (29).

For the ROI-based analyses, in every subject, the responses to each condition were averaged across the voxels within each region. The responses were then averaged across subjects for each region. The responses to the sentences and nonwords conditions were estimated using cross-validation across the functional runs, and therefore, the data used to estimate the effects were always independent of the data used for ROI definition (46). (For example, in a sample scenario of two functional runs, run 1 would be used to define the ROI, and run 2 would be used to estimate the effects. Then, run 2 would be used to define the ROI, and run 1 would be used to estimate the effects. Finally, the estimates would be averaged across the two runs.)

For the whole-brain overlap analyses, we overlaid activation maps (thresholded at $P < 0.001$, uncorrected, or $P < 0.05$, uncorrected, for the liberal thresholding version of the analyses) for (i) the language localizer and (ii) the relevant nonlinguistic task (e.g., Stroop hard > Stroop easy or intact music > scrambled music) in each subject. We then examined the overlap voxels (i.e., voxels responding to both the language localizer and the nonlinguistic task) across subjects to see if any meaningful overlap regions emerge (we used the group-constrained subject-specific method to perform this search for spatially systematic regions) (29). Meaningful overlap regions were defined as regions

that (i) are present in 80% or more of the individual subjects (i.e., have at least one voxel satisfying the predefined criteria, such as a significant response to both the language localizer and the nonlinguistic task, within the borders of the resulting region) and (ii) show replicable effects with cross-validation in a left-out subset of the data.

ACKNOWLEDGMENTS. We thank Josh McDermott for his help during the early stages of developing the language localizer task and in designing and setting up experiment 7 and interpreting its results; Todd Thompson, Jared Novick, and John Gabrieli for their help with pilot studies that motivated experiments 2 and 3; Sam Norman-Haignere for his help in setting up and running the subjects for experiment 7; Alfonso Nieto-Castañón for his help with data analyses; Eyal Dechter, Jason Webster, and Adam Reisner for their help with experimental scripts and running subjects; Christina Triantafyllou, Steve Shannon, and Sheeba Arnold for technical support; two reviewers, Ted Gibson, Julie Golomb, Josh McDermott, Molly Potter, Josh Tenenbaum, and especially, Charles Jennings for comments on the earlier drafts of the paper; and the members of the Kanwisher, Gibson, and Saxe laboratories and John Duncan for helpful discussions of this work. We also would like to acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research, Massachusetts Institute of Technology. This research was funded by Eunice Kennedy Shriver National Institute of Child Health and Human Development Award K99HD-057522 (to E.F.) and an Ellison Foundation grant (to N.K.).

- Broca P (1861) [Notes on the seat of the faculty of articulate language, followed by an observation of aphemia]. *Bull Soc Anat Paris*, 6:330–357. French.
- Wernicke C (1969) *Boston Studies in the Philosophy of Science*, eds Cohen RS, Wartofsky MW (Reidel, Dordrecht, The Netherlands), Vol 4.
- Varley RA, Klessinger NJC, Romanowski CAJ, Siegal M (2005) Agrammatic but numerate. *Proc Natl Acad Sci USA* 102:3519–3524.
- Luria AR, Tsvetkova LS, Futer DS (1965) Aphasia in a composer (V. G. Shebalin). *J Neurol Sci* 2:288–292.
- Peretz I, Coltheart M (2003) Modularity of music processing. *Nat Neurosci* 6:688–691.
- Apperly IA, Samson D, Carroll N, Hussain S, Humphreys G (2006) Intact first- and second-order false belief reasoning in a patient with severely impaired grammar. *Soc Neurosci* 1:334–348.
- Dehaene S, Spelke E, Pinel P, Stanescu R, Tsivkin S (1999) Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science* 284:970–974.
- Novick JM, Trueswell JC, Thompson-Schill SL (2005) Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cogn Affect Behav Neurosci* 5:263–281.
- Levitin DJ, Menon V (2003) Musical structure is processed in "language" areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *Neuroimage* 20: 2142–2152.
- Awh E, et al. (1996) Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psych Science* 7:25–31.
- Badre D, Poldrack RA, Paré-Blagoev EJ, Insler RZ, Wagner AD (2005) Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* 47:907–918.
- Blumstein S (2009) Reflections on the cognitive neuroscience of language. *The Cognitive Neurosciences*, ed Gazzaniga M (MIT Press, Cambridge, MA), pp 1235–1240.
- Braver TS, et al. (1997) A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage* 5:49–62.
- Bunge SA, Klingberg T, Jacobsen RB, Gabrieli JDE (2000) A resource model of the neural basis of executive working memory. *Proc Natl Acad Sci USA* 97:3573–3578.
- Cohen JD, et al. (1994) Activation of the prefrontal cortex in a nonspatial working memory task with functional MRI. *Hum Brain Mapp* 1:293–304.
- Fadiga L, Craighero L, D'Ausilio A (2009) Broca's area in language, action, and music. *Ann N Y Acad Sci* 1169:448–458.
- January D, Trueswell JC, Thompson-Schill SL (2009) Co-localization of stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *J Cogn Neurosci* 21:2434–2444.
- Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. *Science* 302:1181–1185.
- Koelsch S, et al. (2002) Bach speaks: A cortical "language-network" serves the processing of music. *Neuroimage* 17:956–966.
- Levitin DJ, Menon V (2005) The neural locus of temporal structure and expectancies in music: Evidence from functional neuroimaging at 3 Tesla. *Music Percept* 22:563–575.
- Maess B, Koelsch S, Gunter TC, Friederici AD (2001) Musical syntax is processed in Broca's area: An MEG study. *Nat Neurosci* 4:540–545.
- Piazza M, Mechelli A, Price CJ, Butterworth B (2006) Exact and approximate judgements of visual and auditory numerosity: An fMRI study. *Brain Res* 1106:177–188.
- Stanescu-Cosson R, et al. (2000) Understanding dissociations in dyscalculia: A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain* 123:2240–2255.
- Amunts K, et al. (1999) Broca's region revisited: Cytoarchitecture and intersubject variability. *J Comp Neurol* 412:319–341.
- Juch H, Zimine I, Seghier ML, Lazeyras F, Fasel JHD (2005) Anatomical variability of the lateral frontal lobe surface: Implication for intersubject variability in language neuroimaging. *Neuroimage* 24:504–514.
- Tomaiuolo F, et al. (1999) Morphology, morphometry and probability mapping of the pars opercularis of the inferior frontal gyrus: An *in vivo* MRI analysis. *Eur J Neurosci* 11:3033–3046.
- Fedorenko E, Kanwisher N (2009) Neuroimaging of language: Why hasn't a clearer picture emerged? *Lang Linguist Compass* 3:839–865.
- Saxe R, Brett M, Kanwisher N (2006) Divide and conquer: A defense of functional localizers. *Neuroimage* 30:1088–1096.
- Fedorenko E, Hsieh PJ, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N (2010) New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *J Neurophysiol* 104:1177–1194.
- Binder JR, et al. (1997) Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci* 17:353–362.
- Neville HJ, et al. (1998) Cerebral organization for language in deaf and hearing subjects: Biological constraints and effects of experience. *Proc Natl Acad Sci USA* 95: 922–929.
- Braze D, et al. (2011) Unification of sentence processing via ear and eye: An fMRI study. *Cortex* 47:416–431.
- Duncan J, Owen AM (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci* 23:475–483.
- Duncan J (2010) The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends Cogn Sci* 14:172–179.
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202.
- Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36:767–776.
- Amunts K, et al. (2010) Broca's region: Novel organizational principles and multiple receptor mapping. *PLoS Biol* 8:e1000489.
- Petrides M, Pandya DN (1994) Comparative architectonic analysis of the human and the macaque frontal cortex. *Handbook of neuropsychology*, eds Bolter F, Grafman J (Elsevier Science Ltd, Amsterdam), pp 17–58.
- Bush G, Shin LM (2006) The Multi-Source Interference Task: An fMRI task that reliably activates the cingulo-frontal-parietal cognitive/attention network. *Nat Protoc* 1: 308–313.
- Schwarzlose RF, Baker CI, Kanwisher N (2005) Separate face and body selectivity on the fusiform gyrus. *J Neurosci* 25:11055–11059.
- Friederici AD (2002) Towards a neural basis of auditory sentence processing. *Trends Cogn Sci* 6:78–84.
- Hagoort P (2005) On Broca, brain, and binding: A new framework. *Trends Cogn Sci* 9: 416–423.
- Hickok G, Poeppel D (2004) Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 92:67–99.
- Ullman MT (2004) Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92:231–270.
- Kanwisher N (2010) Functional specificity in the human brain: A window into the functional architecture of the mind. *Proc Natl Acad Sci USA* 107:11163–11170.
- Vul E, Kanwisher N (2010) Begging the question: The non-independence error in fMRI data analysis. *Foundational issues for human brain mapping*, eds Hanson S, Bunzl M (MIT Press, Cambridge, MA), pp 71–91.

Supporting Information

Fedorenko et al. 10.1073/pnas.1112937108

SI Text

Details of the Results of the Whole-Brain Overlap Analyses. As stated in the text, when thresholding individual maps at the same level as the level used for the localizer maps in the region of interest (ROI) analyses ($P < 0.001$, uncorrected), no overlap regions emerged as spatially consistent across subjects between language and any of the nonlinguistic tasks. At the liberal ($P < 0.05$, uncorrected) threshold, the following overlap regions emerged (note that, although a liberal threshold is necessary to discover spatially systematic overlap regions, to the extent that the response in these regions replicates in a left-out subset of the data, these regions can be considered meaningful):

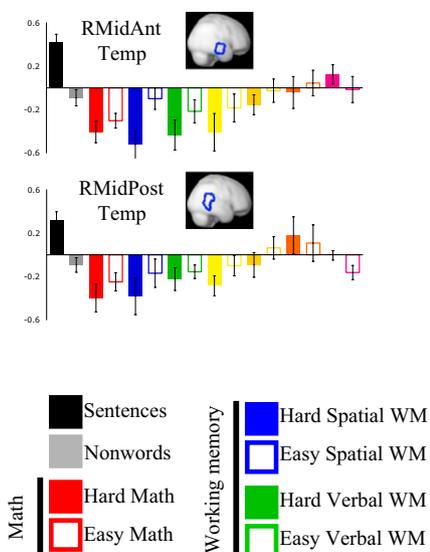
i) A region within our left middle frontal gyrus (LMFG) language ROI that responds to both the language contrast [$t(12) = 5.98, P < 0.0001$] and verbal working memory [WM; t

(12) = 9.38, $P < 0.0001$] (Fig. S2), consistent with a reliable response to verbal WM in the LMFG language ROI.

ii) A region within our LMidPostTemp language ROI that responds to both the language contrast [$t(10) = 8.97, P < 0.0001$] and music [$t(10) = 3.25, P < 0.005$] (Fig. S2), although the response to intact musical stimuli is still weaker than the response to our language control condition (i.e., nonwords), similar to what we found in our ROI-based analyses.

iii) A region in the posterior part of our LIFG language ROI that responds to both the language contrast [$t(10) = 5.14, P < 0.0005$] and verbal WM [$t(10) = 5.57, P < 0.0005$]; a similar region emerges in the overlap analysis of language and Stroop activations, although it is not quite as robust and present in only 71% of the subjects [sentences > nonwords effect, $t(9) = 13.6, P < 0.0001$; hard Stroop > easy Stroop effect, $t(9) = 8.64, P < 0.0001$] (Fig. 4).

Right posterior ROIs



Cerebellar ROIs

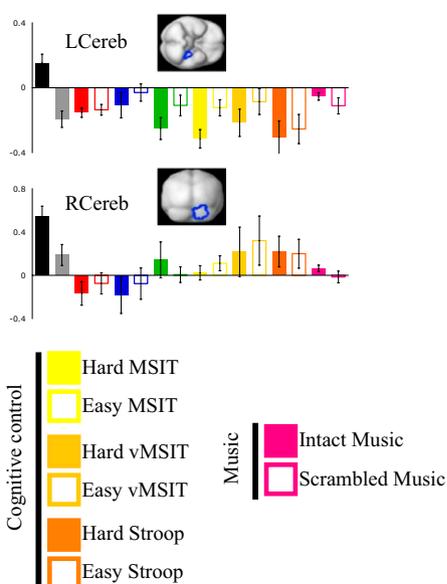
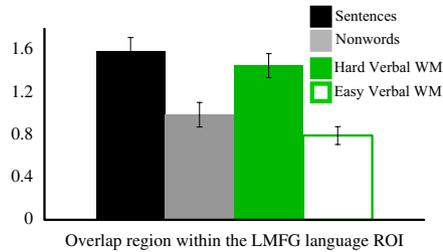
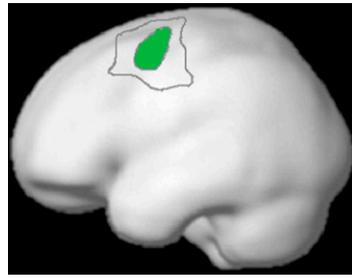


Fig. S1. Responses of the right hemisphere and cerebellar language ROIs to the language and nonlinguistic tasks. Sentences elicit a reliably stronger response than nonwords in each region (t values > 8; P values < 0.0001). However, similar to what we observed for left hemisphere language ROIs, these regions show little response during the nonlinguistic tasks. None of the nonlinguistic task contrasts reach significance, even with no corrections for multiple comparisons (t values < 1.59, P values > 0.07), except for the RMidPostTemp language ROI, which responds more strongly to intact than scrambled musical stimuli [$t(10) = 2.41, P < 0.02$]; this effect, however, does not survive correcting for the number of ROIs.

Language and Verbal WM



Language and Music

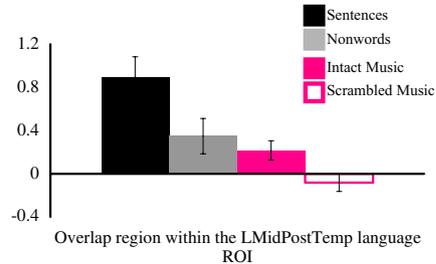
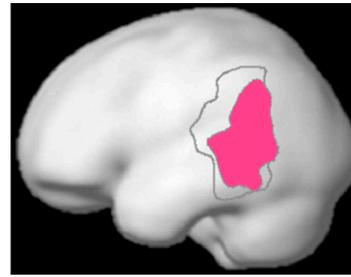


Fig. S2. Overlap regions that emerged in the liberal thresholding ($P < 0.05$, uncorrected) version of the whole-brain overlap analyses (see also Fig. 4). The language and verbal WM overlap region is shown in green (*Left*); this region is present in every subject (average size = 78 voxels) and is contained within the LMFG language ROI (outlined in gray). These data confirm that the LMFG language ROI is not selective for sentence understanding. The language and music overlap region is shown in pink (*Right*); this region is present in 92% of the subjects (average size = 203 voxels) and is contained largely within the LMID-PostTemp language ROI (outlined in gray). The music response in this region is significant but weak, leaving claims about the specificity of this region for high-level linguistic processing largely intact.

Table S1. Statistics associated with Fig. 2: Language, Math, Spatial WM, Verbal WM, and MSIT tasks

Region (average number of voxels; proportion of subjects)	Language (sentences > nonwords)	Math (hard > easy; hard > fix)	Spatial WM (hard > easy; hard > fix)	Verbal WM (hard > easy; hard > fix)	MSIT (hard > easy; hard > fix)
Frontal regions					
LIFGorb (171; 0.89)	$t(42) = 11.9, P < 0.0001$	$t(8) = -1.49, ns; t(8) = -2.83, ns$	$t(9) = -2.33, ns; t(9) = -2.29, ns$	$t(11) = -0.98, ns; t(11) = -3.3, ns$	$t(14) = -2.1, ns; t(14) = -1.6, ns$
LIFG (277; 0.95)	$t(45) = 16.6, P < 0.0001$	$t(10) = 0.01, ns; t(10) = 0.15, ns$	$t(11) = -2.22, ns; t(11) = -1.44, ns$	$t(12) = 1.94, ns (P = 0.038 unc); t(12) = 0.92, ns$	$t(14) = -1.61, ns; t(14) = -2, ns$
LMFG (179; 0.91)	$t(43) = 15.1, P < 0.0001$	$t(10) = 1.01, ns; t(10) = 1.86, ns (P = 0.046 unc)$	$t(11) = 0.05, ns; t(11) = 1.8, ns (P = 0.049 unc)$	$t(12) = 4.31, P < 0.005; t(12) = 4.35, P < 0.005$	$t(14) = -0.58, ns; t(14) = 1.34, ns$
LSFG (187; 0.81)	$t(38) = 11.3, P < 0.0001$	$t(10) = -3.35, ns; t(10) = -6.24, ns$	$t(11) = -3.2, ns; t(11) = -8.23, ns$	$t(11) = -4.93, ns; t(11) = -6.05, ns$	$t(14) = -2.14, ns; t(14) = -0.31, ns$
Posterior regions					
LAnt Temp (172; 1)	$t(47) = 17.6, P < 0.0001$	$t(10) = -1.61, ns; t(10) = -3.99, ns$	$t(11) = -3.34, ns; t(11) = -4.23, ns$	$t(12) = -2.28, ns; t(12) = -5.44, ns$	$t(14) = -2.48, ns; t(14) = -0.78, ns$
LMidAnt Temp (241; 1)	$t(47) = 20.9, P < 0.0001$	$t(10) = -1.63, ns; t(10) = -3.64, ns$	$t(11) = -4.17, ns; t(11) = -4.87, ns$	$t(12) = -3.43, ns; t(12) = -3.68, ns$	$t(14) = -2.84, ns; t(14) = -0.86, ns$
LMidPost Temp (677; 0.98)	$t(46) = 19.7, P < 0.0001$	$t(10) = -0.58, ns; t(10) = -0.9, ns$	$t(11) = -2.78, ns; t(11) = -1.21, ns$	$t(12) = -0.6, ns; t(12) = -0.3, ns$	$t(14) = -2.6, ns; t(14) = -0.52, ns$
LPost Temp (323; 1)	$t(47) = 17.6, P < 0.0001$	$t(10) = -1.45, ns; t(10) = -2.38, ns$	$t(11) = -2.9, ns; t(11) = -1.76, ns$	$t(12) = -1.13, ns; t(12) = -0.62, ns$	$t(14) = -1.97, ns; t(14) = -1.06, ns$
LAngG9 (197; 0.88)	$t(41) = 11.3, P < 0.0001$	$t(10) = -0.18, ns; t(10) = -1.59, ns$	$t(11) = -2.5, ns; t(11) = -1.8, ns$	$t(12) = -3.12, ns; t(12) = -3.02, ns$	$t(14) = -4.12, ns; t(14) = -3.92, ns$

Statistics quantifying the responses of the left-hemisphere language ROIs to the language task (estimated in left-out data) and nonlinguistic tasks, where we include both the critical contrast (hard > easy/intact > scrambled) and the contrast between the hard/intact condition and the fixation baseline for completeness and to aid interpretation. The significance values were Bonferroni-corrected for the number of regions (i.e., multiplied by nine). In cases where the P value reached significance at the $P < 0.05$ level before the Bonferroni correction, we provide these uncorrected (unc) P values in parentheses for completeness. For each region, we report the average size (in voxels; each voxel = 2 mm³) and the proportion of subjects for whom the ROI could be defined. MSIT, multisource interference task.

Table S2. Statistics associated with Fig. 2: vMSIT, Stroop, and Music tasks

Region	vMSIT (hard > easy; hard > fix)	Stroop (hard > easy; hard > fix)	Music (intact > scrambled; intact > fix)
Frontal regions			
LIFGorb (171; 0.89)	$t(11) = -3.9$, ns; $t(11) = -2.69$, ns	$t(13) = -1.37$, ns; $t(13) = 0.43$, ns	$t(11) = 2.27$, ns ($P = 0.022$ unc); $t(11) = 3.23$, $P < 0.05$
LIFG (277; 0.95)	$t(11) = -0.96$, ns; $t(11) = 0.1$, ns	$t(13) = 2.46$, ns ($P = 0.014$ unc); $t(13) = 2.42$, ns ($P = 0.016$ unc)	$t(11) = 1.97$, ns ($P = 0.037$ unc); $t(11) = 3.07$, $P < 0.05$
LMFG (179; 0.91)	$t(11) = -0.22$, ns; $t(11) = 2.54$, ns ($P = 0.014$ unc)	$t(13) = 2.64$, ns ($P = 0.01$ unc); $t(13) = 3.66$, $P < 0.05$	$t(11) = 1.47$, ns; $t(11) = 1.21$, ns
LSFG (187; 0.81)	$t(10) = -2.11$, ns; $t(10) = -0.08$, ns	$t(13) = -2.54$, ns; $t(13) = -0.05$, ns	$t(9) = 1.33$, ns; $t(9) = 1.99$, ns ($P = 0.039$ unc)
Posterior region			
LAntTemp (172; 1)	$t(11) = -0.86$, ns; $t(11) = -0.31$, ns	$t(13) = -0.73$, ns; $t(13) = -1.37$, ns	$t(11) = 2.71$, ns ($P = 0.01$ unc); $t(11) = 1.62$, ns
LMidAntTemp (241; 1)	$t(11) = -3.34$, ns; $t(11) = -1.24$, ns	$t(13) = -1.39$, ns; $t(13) = -1.5$, ns	$t(11) = 2.23$, ns ($P = 0.024$ unc); $t(11) = 1.7$, ns
LMidPostTemp (677; 0.98)	$t(11) = -2.17$, ns; $t(11) = 0.58$, ns	$t(13) = 1.71$, ns; $t(13) = 0.72$, ns	$t(11) = 1.95$, ns ($P = 0.039$, unc); $t(11) = 1.25$, ns
LPostTemp (323; 1)	$t(11) = -0.8$, ns; $t(11) = 1.21$, ns	$t(13) = 0.04$, ns; $t(13) = -1.99$, ns	$t(11) = 0.57$, ns; $t(11) = -1.04$, ns
LAngG (197; 0.88)	$t(11) = -2.74$, ns; $t(11) = -0.91$, ns	$t(13) = -2.9$, ns; $t(13) = -5.76$, ns	$t(11) = 0.52$, ns; $t(11) = -1.32$, ns

vMSIT, verbal multisource interference task; unc, uncorrected.

Table S3. Behavioral data

Task	Accuracies in percent correct		RTs in ms	
	Hard	Easy	Hard	Easy
Expt 1 (math)	83.8 (2.6)	93.7 (1.6)	937 (118)	721 (71)
Expt 2 (spatial WM)	75.7 (1.6)	97.2 (0.9)	1,345 (116)	790 (45)
Expt 3 (verbal WM)	84.3 (2.3)	98.4 (0.4)	2,002 (90)	1,097 (56)
Expt 4 (MSIT)	97.7 (0.7)	99.8 (0.1)	799 (19)	533 (16)
Expt 5 (vMSIT)	92.9 (1.6)	99.2 (0.5)	1,015 (23)	690 (26)

Behavioral results for the nonlinguistic tasks. SEMs are indicated in parentheses (for experiment 3, behavioral responses from one participant were not recorded because of a script error). For experiment 6 (Stroop), in the postscanning questionnaire, participants were asked the following question: "If the easy version of the task (with noncolor adjectives) is zero on a scale from one to five, how difficult is the hard version (with color adjectives)?" The average rating was 2.65 (SE = 0.22). Expt, experiment; MSIT, multisource interference task; vMSIT, verbal multisource interference task.