

Bihemispheric foundations for human speech comprehension

Mirjana Bozic^a, Lorraine K. Tyler^b, David T. Ives^{c,d}, Billi Randall^b, and William D. Marslen-Wilson^{a,1}

^aMedical Research Council Cognition and Brain Sciences Unit, Cambridge CB2 7EF, United Kingdom; ^bDepartment of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, United Kingdom; ^cDepartment of Physiology, Development, and Neuroscience, University of Cambridge, Cambridge CB2 3EG, United Kingdom; and ^dLaboratoire de Psychologie de la Perception, Centre National de la Recherche Scientifique, Université Paris Descartes, 75005 Paris, France

Edited by Willem J. M. Levelt, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, and approved August 17, 2010 (received for review January 14, 2010)

Emerging evidence from neuroimaging and neuropsychology suggests that human speech comprehension engages two types of neurocognitive processes: a distributed bilateral system underpinning general perceptual and cognitive processing, viewed as neurobiologically primary, and a more specialized left hemisphere system supporting key grammatical language functions, likely to be specific to humans. To test these hypotheses directly we covaried increases in the nonlinguistic complexity of spoken words [presence or absence of an embedded stem, e.g., *claim (clay)*] with variations in their linguistic complexity (presence of inflectional affixes, e.g., *play+ed*). Nonlinguistic complexity, generated by the on-line competition between the full word and its onset-embedded stem, was found to activate both right and left fronto-temporal brain regions, including bilateral BA45 and -47. Linguistic complexity activated left-lateralized inferior frontal areas only, primarily in BA45. This contrast reflects a differentiation between the functional roles of a bilateral system, which supports the basic mapping from sound to lexical meaning, and a language-specific left-lateralized system that supports core decompositional and combinatorial processes invoked by linguistically complex inputs. These differences can be related to the neurobiological foundations of human language and underline the importance of bihemispheric systems in supporting the dynamic processing and interpretation of spoken inputs.

brain | language | morphology | inflection

Considered as a neuroscientific phenomenon, human language comprehension—and human language function in general—is both remarkably specific and remarkably distributed in its neural instantiation. One hundred fifty years of neuropsychological research demonstrate the irreplaceable role of left hemisphere perisylvian networks (linking left inferior frontal and posterior temporal brain areas) in supporting key combinatorial language functions in the domain of inflectional morphology and syntax. At the same time, evidence from lesion studies and neuroimaging of the intact brain shows that dynamic access to lexical meaning, and the ability to rapidly construct semantic and pragmatic interpretations of incoming speech, can remain strikingly intact even when the left perisylvian language areas are destroyed, implying a distributed bihemispheric substrate. We believe that these contrasts reflect a fundamental distinction in the types of processing operations underlying normal language comprehension and in the neural networks that support these functions. In this event-related fMRI study of lexical processing, we test the hypothesis that general purpose processing demands engage both right and left perisylvian systems, whereas linguistically specific processing demands selectively engage left hemisphere systems.

A critical element of this account is its emphasis on the bihemispheric foundations of human speech communication. Functional imaging shows that bilateral activation of temporal lobe structures in and around primary auditory processing areas is not limited to low-level acoustic and phonetic analyses, but also implicates lexical processes. There is complementary evidence from neuropsychological and neuroimaging studies for the

involvement of bihemispheric temporal regions in processes related to accessing lexical meaning (1). Compared with low-level acoustic baselines, spoken words activate superior and middle temporal gyri bilaterally, regions related to meaning-based processes in language comprehension. Although researchers disagree on the aspects of lexical processing supported by left and right hemispheres (e.g., ref. 2), neuropsychological data strongly suggest that both hemispheres have access to representations of lexical meaning. Patients with extensive damage to left frontal and superior temporal regions but spared right hemisphere (RH) homologs can still recognize simple spoken words and produce semantic priming effects and reaction times within the normal range, demonstrating rapid and efficient access to lexical identity and meaning from auditory inputs (3).

At the same time, superimposed on this highly effective basic bilateral system is a left-dominant perisylvian core language system, linking left inferior frontal cortex (LIFC) with temporal and parietal cortices. Damage to these regions can cause permanent disruption of language functions although damage to parallel RH regions typically does not. Particularly salient here is disruption to the combinatorial aspects of language. Patients with left hemisphere (LH) damage, especially in LIFC, have problems with syntax and inflectional morphology, both in production and in comprehension (4). Evidence from neuroimaging confirms the salience of LH contributions to language, primarily located in inferior frontal, temporal, and inferior parietal cortex (5). In terms of grammatical function, research focused around regular inflectional morphology reveals a key role for this LH fronto-temporal network in processes related to morpho-phonology and morpho-syntax. The selective deficits for regular inflectional morphology, shown by many LH nonfluent patients, point to a decompositional network linking LIFC with superior and middle temporal cortex (6). This network handles the processing of regularly inflected words (such as *joined* or *treats*), which are argued not to be stored as whole forms and instead require morpho-phonological parsing to be segmented into stems and inflectional affixes.

These observations, combining detailed psycholinguistic and neural theories, raise basic questions about the properties of the processing systems that underpin human language function. To understand what is special about the core LH fronto-temporal system, we need a better understanding both of the properties of RH systems and of the aspects of language processing that are

Author contributions: M.B., L.K.T., B.R., and W.D.M.-W. designed research; M.B. performed research; D.T.I. and B.R. contributed new reagents/analytic tools; M.B., L.K.T., and W.D.M.-W. analyzed data; and M.B., L.K.T., and W.D.M.-W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The fMRI data have been deposited in XNAT Central, <http://central.xnat.org> (Project ID: speech).

¹To whom correspondence should be addressed. E-mail: w.marslen-wilson@mrc-cbu.cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1000531107/-DCSupplemental.

common to both hemispheres. These considerations lead to specific questions about the processing and representational properties of the basic bilateral system hypothesized to support the access of simple spoken words. We predict that both hemispheres will be involved to a similar degree when the linguistic input consists of morphologically simple, high-frequency, high-imageability words. These types of words, in a language like English, will engage general purpose perceptual processing mechanisms that are not specifically linguistic in nature. Such mechanisms are engaged by the unfolding nature of the speech signal, where word onsets map onto multiple lexical representations. This triggers competition between these perceptual alternatives and requires additional processing to select the correct candidate. Competition and selection between multiple alternatives are common to a range of cognitive functions, from visual perception to working memory, and are associated with increased bilateral frontal activation (7).

To selectively engage these mechanisms we used words that are not linguistically complex and do not require decompositional analysis, but are nonetheless perceptually complex. These are morphologically simple words like *lawn* and *ramp*, which can be directly mapped onto underlying full-form representations, but that also have an onset-embedded stem (*law/lawn, ram/ramp*). The presence of this second potential candidate for lexical access creates competition between the two simultaneously activated stems. Successful choice of the correct lexical representation will therefore involve additional decision and selection processes. Previous language research indicated that selection processes require the involvement of inferior frontal regions (8), but with few exceptions (e.g., ref. 9) focused on LH effects. On the approach taken here, selection processes operate on lexical candidates whose analysis is conducted bilaterally, triggering the involvement of bilateral IFC.

When the input changes to include linguistically more complex material, we predicted a shift in fronto-temporal involvement across the hemispheres. This shift affects both the degree of left (L) temporal involvement in superior temporal and middle temporal gyri (STG/MTG) and the degree of LH fronto-temporal interaction. Here we focused on regular inflectional morphemes, such as the English past tense {-d}, which lead a spoken lexical input to interact differentially with the machinery of lexical access and linguistic interpretation (10). When a phonetic input such as /preId/ is encountered, corresponding to the word *prayed*, the listener must both access the lexical information associated with the stem {*pray*} and extract the processing implications of the presence of the grammatical morpheme {-d}. It is the capacity to support this segmentation into stem- and affix-based processes that is disrupted in patients with LH perisylvian damage (6). A key concept here is the notion of the *inflectional rhyme pattern* (IRP) as a dynamic modulator of fronto-temporal integration that indicates whether the current speech input carries inflectional information necessary for its on-line structural interpretation (11).

The IRP is a phonological pattern in English that signals that the ending of a complex word may be an inflectional affix and therefore should not be treated as part of the stem. It is defined in terms of specific phonological markers—whether the final consonant is coronal [i.e., (d, t, s, z)] and whether it agrees in voice with the preceding segment—which are shared by all regular {-d} and {-s} inflections in English. It may be present both in real inflected forms (*bowled, packs*), in pseudoinflected forms where a simple word has an apparent inflectional ending (*mold, flax*), and in nonword forms (*nold, dax*). Recent fMRI results suggest that the presence of the IRP has an across-the-board triggering effect on LIFC activation, correlated with L temporal lobe and L anterior cingulate activation (11). Consistent with this, patients with LIFC damage show marked disruptions in performance when confronted with IRP stimuli, again independent of lexical status (12). Behavioral studies with unimpaired young adults (13) show increased processing complexity associated with the presence of the IRP for real words and nonwords and for both major types of regular inflection in English (the past tense -d and plural -s).

These results indicate a dynamic interaction between stem-access processes mediated bilaterally by temporal lobe structures and

predominantly left frontal systems critical for morpho-phonological and syntactic parsing of the input stream (6, 12). To probe these interactions we covaried the two sources of processing complexity described above: whether a word ends in the IRP, invoking specifically linguistic complexity, and whether it has an embedded stem, modulating general processing complexity. Simple words like *dream*, exhibiting neither type of complexity, contrast both with words like *claim*, which have an embedded stem (*clay*) but no IRP, and with words like *trend*, which end in a potential suffix, but have no embedded stem. Intermediate cases like *played* and *trade* exhibit both forms of complexity—each ends in a potential inflectional affix, and each has a potential embedded stem (*play* and *tray*). Variations in the degree of stem competition should modulate activation bilaterally, with greater competition increasing the likelihood of IFC involvement. In contrast, presence or absence of the IRP should primarily modulate LH activation, especially where fronto-temporal links are concerned.

Because this approach predicts that LH networks, especially LIFC, should be sensitive to both of these types of processing complexity, it raises the question of whether the activations elicited will be spatio-temporally distinct or overlapping. Issues of functional specialization in LIFC are widely debated and allow for several possible outcomes: If linguistic computations (tapped into by the IRP conditions) are domain and region specific, they may generate activity in more dorsal LIFC (BA44/45), whereas domain-general computations, elicited by stem competition, should engage more orbito-frontal areas, including BA47 (14, 15). Alternatively, on a network-based account (16), the same LIFC areas may support both linguistic and nonlinguistic processing functions, but do so in the context of different distributed processing networks.

Finally, to be able to isolate the lexical processing network from lower-level auditory processing, we compared word processing against a baseline that unambiguously separates complex auditory processing from specifically acoustic-phonetic analyses. Correct choice of baseline is critical for a convincing demonstration of RH lexical processing, where inappropriate baselines lead either to false negatives or to false positives. Our baseline for speech sounds was “musical rain” (MuR) as described in ref. 17. MuR was chosen over acoustic baselines such as spectrally rotated, noise-vocoded or reversed speech because it provides a principled basis for claiming that it closely tracks the acoustic properties of speech, while at the same time not being interpretable as speech. The temporal envelope of each MuR segment was extracted from the corresponding speech segment, and the two were matched in terms of the long-term spectro-temporal distribution of energy and root mean square (RMS) level. Nevertheless, despite the similarities in the distribution of energy over frequency and time, the absence of continuous formants in the signal means that MuR is not interpretable as being the output of a human vocal tract and is not heard as speechlike. MuR produces a similar level of neural activation to that of vowels in the auditory pathway up to and including primary auditory cortex in Heschl’s gyrus and planum temporale (17). In secondary auditory regions, however, such as anterior superior temporal cortex, it produces much less activation than the corresponding speech.

In the context of this strict auditory baseline, we can evaluate directly the hemispheric distribution of contrasting lexical processing systems, using a set of stimuli that covary stem competition and suffix decomposition demands (Table 1). Words in the regular past tense condition (e.g., *prayed*) were all morphologically complex, consisting of stems and suffixes. These words should trigger a decomposition process on the basis of the presence of the suffix. Although they have a phonologically embedded stem, this may not trigger strong lexical competition because *prayed* is not thought to be separately lexically represented (6). The second condition consisted of monomorphemic words with a real word embedded at the onset and an IRP ending, e.g., *trade*. Words from this condition can be processed as *tray+ed*, thus triggering both stem competition and suffix decomposition. The third condition consisted of monomorphemic words with an IRP ending but without an onset embedded word, e.g., *blend*. Here, the presence of the IRP ending can

Table 1. Experimental conditions and behavioral results in milliseconds

Experimental condition	Example	Embedded stem	Suffix (IRP)	RT (SD)*
1. Regular past tense	<i>Prayed</i>	? (pray)	Y	970 (90)
2. Pseudoregulars	<i>Trade</i>	Y (tray)	Y	942 (65)
3. No stem, with IRP	<i>Blend</i>	N	Y	955 (68)
4. Stem only	<i>Claim</i>	Y (clay)	N	967 (92)
5. Simple	<i>Dream</i>	N	N	924 (71)

N, no; Y, yes.

*Behavioral results, reaction time.

generate a decomposition process, but there is no basis for stem competition because the remaining set of phonemes (e.g., *blen*) does not form a real word. Words in the fourth condition were monomorphemic, with an embedded stem but no IRP ending, e.g., *claim*. These words should trigger stem competition because of the embedded stem (e.g., *clay*). The fifth condition consisted of monomorphemic words with neither embedded stems nor IRPs, e.g., *dream*. These are simple words that do not exhibit either linguistic or perceptual processing complexity, placing least demands on both bilateral and L perisylvian processing systems.

Results

Behavioral Data. Participants performed a gap detection task (18). All errors (1.5%) and time-outs [reaction time (RT) >3,000 ms, 0.1%] were removed, and data were inverse transformed. Mean reaction times and SD for the test conditions (“no” responses in the gap detection task) are shown in Table 1.

A univariate ANCOVA was performed on the inverse-transformed RTs, with condition as a fixed factor and speech file duration as a covariate. The results showed a significant difference in RTs between conditions [$F_1(4, 44) = 11.6, P < 0.01$; $F_2(4, 194) = 2.6, P < 0.05$]. Effects of IRP presence and embeddedness were tested by grouping conditions according to presence of IRP and embedded stems and comparing them against the RTs for simple words. Simple words (e.g., *dream*) were responded to significantly faster than words with IRPs (*prayed, trade, blend*) [$F_1(3, 33) = 10.5, P < 0.01$; $F_2(1, 159) = 6.9, P < 0.01$], or embedded stems (*trade, claim*) [$F_1(2, 22) = 18.2, P < 0.01$; $F_2(1, 119) = 5.7, P < 0.05$]. There was no significant difference between RTs for the words with IRPs and embedded stems [$F_1(3, 33) = 5.2, P < 0.01$; $F_2(3, 159) = 0.83, P > 0.1$]. These results indicate a general increase in lexical processing load for the complex items, consistent with earlier studies using gap detection (18).

Imaging Data. On the basis of previous evidence (1, 15) and our predictions, we focused on bilateral fronto-temporo-parietal regions as the volume of interest for the analyses. Using WFU Pickatlas, a mask was constructed, consisting of bilateral temporal lobes (superior, middle and inferior temporal gyri, and angular gyrus), inferior frontal gyri (pars orbitalis, pars opercularis, pars triangularis, and precentral gyrus), insula, and the anterior cingulate. We first established the network that supports complex acoustic processing by subtracting null events from MuR. This contrast showed bilateral activity in primary auditory cortices and surrounding superior temporal regions (BA42/22, peaks at $-66 -30 14$ and $68 -34 14$), consistent with previous results (5, 17) (Fig. 1 and *SI Appendix, Table S1*). Laterality analyses showed generally bilateral activation, with only medial temporal activation stronger in LH compared with RH. To conduct these laterality analyses, EPI images were normalized onto a symmetrical T1 template, and the standard first-level analysis was performed. The resulting maps were flipped along the y axis and compared with the original maps in a series of t tests (19).

To extract the activity specifically related to speech-driven lexical processing we contrasted words with the MuR baseline. This comparison showed that lexical processes activated regions that

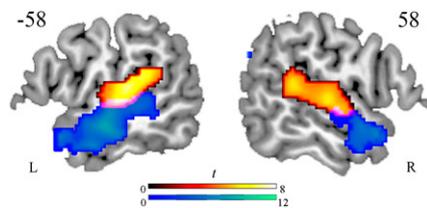


Fig. 1. Significant activation for complex acoustic processing (orange) and linguistic processing (blue) rendered onto the surface of a canonical brain. Threshold at $P < 0.001$ voxel level and $P < 0.05$ cluster level corrected for multiple comparisons is shown.

are anterior and ventro-lateral to the activity observed for lower-level auditory processing (Fig. 1). These lexical activations were in bilateral middle temporal gyri (BA21) extending into fusiform and hippocampal regions (BA20/37), angular gyrus (BA39), and anterior cingulate (BA32), as well as L inferior frontal regions on the lower statistical threshold of 0.01. Despite the significant RH activity, these activations had a strong LH bias: The laterality analyses showed that lexical activity (words minus MuR) was significantly stronger in LIFC (BA44/45/47) and left angular gyrus (BA39), whereas no region showed significantly more RH lexical activation (*SI Appendix, Table S2*).

Turning to our key questions concerning patterns of fronto-temporal activation as a function of different types of lexical processing complexity, and their hemispheric distribution, we first conducted exploratory analyses using the multivariate linear approach implemented in the MLM toolbox (20). These analyses classified conditions on the basis of the patterns of activation they evoke in the fronto-temporal volume of interest (relative to the MuR baseline) and calculated orthogonal eigenvectors that show greatest difference between conditions. Two eigenvectors speak directly to the critical contrasts in this study (details in *SI Appendix*). The first of these (Fig. 2A) dissociates conditions with and without an embedded stem, with the divergence between *claim, trade, and prayed* (each with actual or potential embedded stems) versus *blend and dream* (no embedded stems). Overlaid on the brain (Fig. 2C) this eigenvector component shows a bilateral fronto-temporal distribution with IFC activation focused in BA47 (pars orbitalis) and extending dorsally into BA45 (pars triangularis). In separate analyses on RIFC and LIFC regions of interest (ROIs), the same dominant pattern emerges, accounting for 77% of the variance in RIFC and 64% of the variance in LIFC. The second eigenvector of interest (Fig. 2B) dissociates conditions with

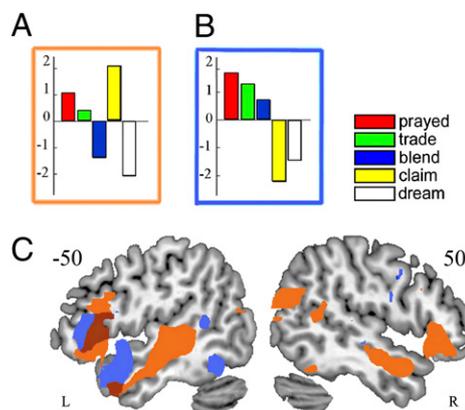


Fig. 2. Selected MLM eigenvectors and their overlay. Bar graphs are in the unit of the beta images. (A) Component dissociating embedded from nonembedded words; (B) component dissociating IRP from non-IRP words; (C) overlay on a canonical brain of the embeddedness (orange) and IRP (blue) components, shown at $t > 1$, with the overlap between them in purple.

introduced, the resulting activation shifted to a strongly left-lateralized pattern of IFC activity (BA45). These peaks of activity in prefrontal sites extend into more posterior temporal regions at a lower threshold, consistent with the view that activity in LIFC and RIFC reflects membership of broader fronto-temporal networks, as indicated in earlier research (e.g., refs. 7, 12, 16, 21). Parallel results emerged from the illustrative multivariate analyses (Fig. 2): A component that dissociates words that do and do not have an embedded stem showed a robust bilateral fronto-temporal distribution, whereas an orthogonal component that dissociated words with and without a potential past tense ending showed a primarily left-lateralized activation pattern.

This set of results enables us to pull together a range of existing but fragmented evidence about the functional architecture of the human speech comprehension system. Complementary results from neuropsychological and neuroimaging research suggest that dynamic speech comprehension processes rely on a distinctly bilateral neural network (1, 3, 16). In healthy volunteers activation for words against a low-level acoustic baseline is typically observed in superior and middle temporal regions (2, 5), with the exact localization determined by complexity of the baseline stimulus and task demands (22). Temporal lobe structures have long been implicated in lexical-semantic processes (23), even though authors disagree on the precise role of different temporal lobe regions in the process of sound-to-meaning mapping. The bilateral foundation of the speech comprehension system is further supported by neuropsychological data showing that patient performance on linguistically simple spoken words remains preserved even when the LH perisylvian areas are destroyed (3, 16). In addition to the involvement of bilateral MTG, we also observed activation in bilateral angular gyri, thought to be involved in the mapping between spoken forms and their meanings (11), medial temporal lobe structures, required for the encoding and maintenance of verbal material (24), and anterior cingulate, implicated in modulating fronto-temporal integration in lexical context (21).

Although researchers generally agree on the role of temporal regions in speech comprehension, there is much less agreement on the functional role of inferior frontal regions and how they interact during spoken language comprehension. The focus has typically been on LIFC, which has been implicated in language-specific functions such as phonological analysis, syntactic parsing, and morphological segmentation, as well as wider, non-language-specific processes such as memory retrieval or selection between competing alternatives (8, 14). A variety of different mechanisms have been proposed for these different functions. The most salient contrast is between functional modularity accounts, where specific areas within the LIFC perform unique language-specific functions, and network-based accounts, where the processing capacities of a given region are recruited to support both domain-specific and domain-general functions, as determined by the properties of the different distributed networks to which these regions are linked and are differentially activated by specific inputs.

Our results, although consistent with earlier claims that different regions of LIFC support different aspects of language-related function, are clearly inconsistent with the claim that this is based on functional modularity. We saw some spatial dissociation between linguistic IRP-related activations and general processing activations, in the sense that both IRP and embeddedness contrasts activated BA45, but only the embeddedness contrast extended to the more ventral BA47. This distribution of effects is broadly consistent with proposals that anterior ventral regions (BA47/45) support processing of lexical semantics, whereas more posterior dorsal areas (BA45/44) support syntactic processing (14)—although it is important to note that the contrasts in this experiment did not directly involve the phrasal and sentential aspects of syntactic and semantic combination. However, as is made clear in *SI Appendix, Fig. S4*, the IRP-related activations in L BA45 seem to fully overlap with those related to embeddedness. This overlap of linguistic and nonlinguistic operations suggests that the functional role and the degree of left-lateralization of fronto-temporal interaction patterns will vary depending on the nature of the speech input. Lexical

processing demands that are not specifically linguistic can clearly activate inferior frontal areas bilaterally.

Competition and selection in speech comprehension reflects the sequential nature of the speech signal. As acoustic information unfolds over time, it triggers the simultaneous activation of multiple competing lexical candidates (25). Successful comprehension requires selection of the correct lexical candidate, and although we focused on lexical selection here, it is unlikely that the underlying processes are language specific. Instead, they reflect an increase in processing demands due to competing perceptual alternatives, general across a wide range of cognitive functions and commonly associated with increased activation in bilateral inferior frontal regions (7). These frontal areas may exert top-down control over posterior regions that store representations and mediate access to these representations from spoken inputs, by biasing the network toward the correct interpretation and supporting information retrieval in ambiguous contexts. Within the language domain, recent studies show that semantically ambiguous words and pairs of near-synonyms, which can map onto multiple competing representations in semantic space, trigger a bilateral increase in activation in BA45/47 (8, 26). This pattern of results is consistent with the effects we observed here: Compared with simple words, words with embedded stems activate bilateral frontal regions (BA45/47), and this pattern was dissociable from the left-lateralized activation due to the presence of linguistic complexity.

In sum, the results of the current study and existing evidence about the neuro-functional properties of the speech comprehension network indicate a functional dissociation between bilateral and left-lateralized fronto-temporal subsystems. A neurobiological context for this duality of organization is suggested by recent research into the neural systems underlying complex auditory object processing (including conspecific vocal calls) in nonhuman primates (27). Basic similarities have emerged between the functional architecture of the macaque and the human comprehension systems. Following bilateral input to primary auditory cortex, processing streams extend ventrally and laterally in a hierarchical manner, projecting to frontal, parietal, and posterior temporal regions. Species-specific calls in the macaque activate L and R fronto-temporal regions (homologs of Broca's and Wernicke's areas), possibly supporting the interpretation of vocal calls in their situational and social contexts. These observations point to a bihemispheric substrate for the processing and interpretation of complex auditory signals that remains fundamental to human speech comprehension as well.

The critical difference from these primate systems is the human-specific LH specialization that supports the core morpho-syntactic properties of language. These grammatical capacities are uniquely dependent on the left fronto-temporal system, and RH fronto-temporal regions cannot take over these functions. Recent research comparing white matter connections between frontal and temporal regions in humans, macaques, and chimpanzees (28) shows that major evolutionary changes have taken place in this LH fronto-temporal network. Comparing humans with chimpanzees, and chimpanzees with macaques, there is a striking, hemispherically asymmetrical increase in the LH arcuate fasciculus that links posterior temporal and inferior frontal areas critical for human language. This emerging LH perisylvian system, however, should be seen as continuing to function in the broader context of the bihemispheric systems for processing and interpreting speech inputs that have emerged so clearly in the current study.

Methods

Participants. Twelve right-handed (seven males) native speakers of British English participated in the study. They had normal or corrected-to-normal vision and had been screened for neurological or developmental disorders. All gave informed consent and were paid for their participation. The study was approved by the Peterborough and Fenland Ethical Committee.

Experimental Design. Stimuli. There were five test conditions with 40 words each (Table 1). All words were matched across conditions on word and

lemma frequency, familiarity, imageability sound file length, and cohort-related variables (all $P > 0.1$, based on CELEX and MRC Psycholinguistic databases; see *SI Appendix* for details). The 200 test words were mixed with 200 filler words, 200 acoustic baseline trials, and 160 silence trials. Filler words were a mix of simple words (two-thirds) and words with embedded stems or suffixes (one-third), to balance out the structure of the test items. The acoustic baseline trials were constructed to share the complex auditory properties of speech but not trigger phonetic interpretation. They were produced by applying a temporal envelope, extracted for each of the speech tokens, to MuR to produce envelope-modulated MuR tokens. The procedure for generating musical rain itself is described in ref. 17 and summarized in *SI Appendix*. The technique produces nonspeech auditory stimuli in which the long-term spectro-temporal distribution of energy is closely matched to that of the corresponding speech stimuli (*SI Appendix, Fig. S1*).

Procedure. The design of the experiment required a task that engages lexical processing but allows us to keep the task requirements constant across speech and nonspeech stimuli. To this end we used gap detection, a nonlinguistic task known to engage lexical processing on line (18). Short silent gaps (400 ms) were inserted in 25% of trials (100 filler words and 50 MuR trials) and participants were required to decide as quickly and accurately as possible whether words and MuR sounds contained a silent gap. For “no” responses participants pressed the button under their index finger and for “yes” responses the button under their middle finger. Only gap-absent trials were subsequently analyzed.

The words were recorded in a sound-proof room by a female native speaker of British English onto a DAT recorder, digitized at a sampling rate of 22 kHz with 16-bit conversion, and stored as separate files using CoolEdit. CoolEdit was also used for gap insertion. Items were presented using in-house software and participants heard the stimuli binaurally over Etymotic R-30 plastic tube phones. There were a total of 760 trials, pseudorandomized with respect to their type (test, filler, baseline, null) and presence or absence of gaps and presented in four blocks of 190 items (10.7 min) each. There were 5 items at the beginning of each block to allow the signal to reach equilibrium.

The experiment started with a short practice session outside the scanner, where participants were given feedback on their performance. Participants were instructed to keep their eyes closed during the scanning.

Scanning was performed on a 3T Trio Siemens Scanner at the Medical Research Council Cognition and Brain Sciences Unit (MRC-CBU), University of Cambridge, using a fastsparse imaging protocol to minimize the interference of scanner noise (gradient-echo EPI sequence, repetition time [TR] = 3.4 s, acquisition time [TA] = 2 s, echo delay time [TE] = 30 ms, flip angle 78, matrix size 64×64 , field of view [FOV] = 192×192 mm, 32 oblique slices 3 mm thick, 0.75 mm gap). MPRAGE T1-weighted scans were acquired for anatomical localization. Stimuli were onset-jittered relative to scan offset by 100–300 ms to allow more effective sampling. Imaging data were preprocessed and analyzed in SPM5. Preprocessing was performed using Automated Analyses Software (MRC-CBU) and involved image realignment to correct for movement, segmentation, and spatial normalization of functional images to the MNI reference brain and smoothing with a 10-mm isotropic Gaussian kernel. The data for each subject were analyzed using the general linear model, with four blocks and 10 event types (five conditions, fillers, MuRs, gap words, gap MuRs, and null). The neural response for each event type was modeled with the canonical haemodynamic response function (HRF) and its temporal derivative. Motion regressors were included to code for the movement effects. Sound length was entered as a parametric modulator for words and musical rain sounds without embedded gaps. A high-pass filter with a 128-s cutoff was applied to remove low-frequency noise. Contrast images were combined into a group random effects analysis, and results were thresholded at uncorrected voxel level of $P < 0.001$ and cluster level of $P < 0.05$ corrected for multiple comparisons.

ACKNOWLEDGMENTS. We thank Carolyn McGettigan for her help with stimulus preparation. This research was supported by Medical Research Council Cognition and Brain Sciences Unit funding (U.1055.04.002.00001.01) (to W.D.M.-W.), a Medical Research Council program grant (to L.K.T.), and a European Research Council Advanced Grant (to W.D.M.-W.).

- Jung-Beeman M (2005) Bilateral brain processes for comprehending natural language. *Trends Cogn Sci* 9:512–518.
- Hickok G, Poeppel D (2000) Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci* 4:131–138.
- Longworth CE, Marslen-Wilson WD, Randall B, Tyler LK (2005) Getting to the meaning of the regular past tense: Evidence from neuropsychology. *J Cogn Neurosci* 17:1087–1097.
- Goodglass H, Christiansen JA, Gallagher R (1993) Comparison of morphology and syntax in free narrative and structured tests: Fluent vs. nonfluent aphasics. *Cortex* 29:377–407.
- Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26:100–107.
- Marslen-Wilson WD, Tyler LK (2007) Morphology, language and the brain: The decompositional substrate for language comprehension. *Philos Trans R Soc Lond B Biol Sci* 362:823–836.
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202.
- Thompson-Schill SL, D’Esposito M, Kan IP (1999) Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron* 23:513–522.
- Raposo A, Moss HE, Stamatakis EA, Tyler LK (2006) Repetition suppression and semantic enhancement: An investigation of the neural correlates of priming. *Neuropsychologia* 44:2284–2295.
- Marslen-Wilson WD, Tyler LK (1998) Rules, representations, and the English past tense. *Trends Cogn Sci* 2:428–435.
- Tyler LK, Stamatakis EA, Post B, Randall B, Marslen-Wilson WD (2005) Temporal and frontal systems in speech comprehension: An fMRI study of past tense processing. *Neuropsychologia* 43:1963–1974.
- Tyler LK, Randall B, Marslen-Wilson WD (2002) Phonology and neuropsychology of the English past tense. *Neuropsychologia* 40:1154–1166.
- Post B, Marslen-Wilson WD, Randall B, Tyler LK (2008) The processing of English regular inflections: Phonological cues to morphological structure. *Cognition* 109:1–17.
- Hagoort P (2005) On Broca, brain, and binding: A new framework. *Trends Cogn Sci* 9:416–423.
- Friederici AD, Rüschemeyer S-A, Hahne A, Fiebach CJ (2003) The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cereb Cortex* 13:170–177.
- Tyler LK, Marslen-Wilson WD (2008) Fronto-temporal brain systems supporting spoken language comprehension. *Philos Trans R Soc Lond B Biol Sci* 363:1037–1054.
- Uppenkamp S, Johnsrude IS, Norris D, Marslen-Wilson WD, Patterson RD (2006) Locating the initial stages of speech-sound processing in human temporal cortex. *Neuroimage* 31:1284–1296.
- Mattys SL, Clark JH (2002) Lexical activity in speech processing: Evidence from pause detection. *J Mem Lang* 47:343–359.
- Liégeois F, et al. (2002) A direct test for lateralization of language activation using fMRI: Comparison with invasive assessments in children with epilepsy. *Neuroimage* 17:1861–1867.
- Kherif F, et al. (2002) Multivariate model specification for fMRI data. *Neuroimage* 16:1068–1083.
- Stamatakis EA, Marslen-Wilson WD, Tyler LK, Fletcher PC (2005) Cingulate control of fronto-temporal integration reflects linguistic demands: A three-way interaction in functional connectivity. *Neuroimage* 28:115–121.
- Giraud AL, et al. (2004) Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cereb Cortex* 14:247–255.
- Hickok G, Poeppel D (2004) Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 92:67–99.
- Strange BA, Otten LJ, Josephs O, Rugg MD, Dolan RJ (2002) Dissociable human perirhinal, hippocampal, and parahippocampal roles during verbal encoding. *J Neurosci* 22:523–528.
- Marslen-Wilson WD, Welsh A (1978) Processing interactions and lexical access during word recognition in continuous speech. *Cogn Psychol* 10:29–63.
- Rodd JM, Davis MH, Johnsrude IS (2005) The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb Cortex* 15:1261–1269.
- Gil-da-Costa R, et al. (2006) Species-specific calls activate homologs of Broca’s and Wernicke’s areas in the macaque. *Nat Neurosci* 9:1064–1070.
- Rilling JK, et al. (2008) The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat Neurosci* 11:426–428.

Supporting Appendix

PNAS MS 2010-00531R: “ Bi-hemispheric foundations for human speech comprehension”

Contents:

1. Musical Rain
2. MLM analyses
3. Univariate Analyses: Lower thresholds
4. Overlap in LIFG activations
5. Analyses of cohort effects and non-initial embeddedness
6. Table S1
7. Table S2
8. Table S3
9. Table S4
10. Experimental stimuli

1. Musical Rain

Musical Rain – method of production.

The production of musical rain (17) uses a technique similar to that used to generate synthetic vowel sounds, by summing together continuous streams of four damped sinusoids. Typically the frequency of the sinusoids represents the formant frequencies and the repetition rate represents the pitch of the vowel. However, for the production of MuR, the carrier frequencies and the repetition rates of each of these sinusoids are randomized: over about a one octave range for the carrier frequency, and over a 20-ms period for the repetition rate. Temporal envelopes are extracted by taking the absolute value of the Hilbert transform for each speech token, and smoothed by low-pass filtering at 20 Hz. The temporal envelopes are then used to modulate the envelopes of the corresponding MuR tokens. The RMS level of the MuR is also matched to that of the speech, producing stimuli in which the long-term spectro-temporal distribution of energy is matched to that of the corresponding speech stimuli (Figure S1).

Figure S1

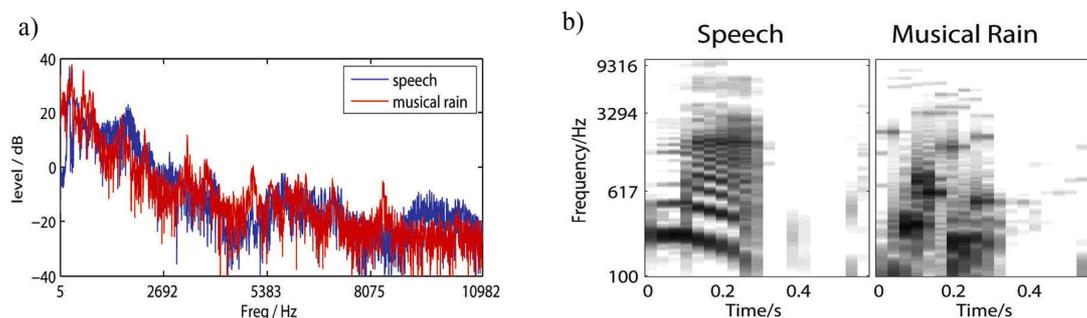


Fig S1: Properties of Musical Rain (MuR tokens and matched speech samples)

(a) Fourier transforms of a speech token and the corresponding Musical Rain (MuR) sample. The similarity of the two Fourier transforms shows that the frequency spectra are well matched. (b) Spectrograms of the same tokens of speech and MuR, showing how modulating the envelope of the MuR produces similar patterns in the distribution of energy over the duration of the sound. The formants are clearly visible in the speech spectrogram but are absent in the MuR spectrogram.

2. MLM analyses

MLM analysis procedures

The MLM procedure used here is derived from the canonical variate analysis (CVA), a multivariate technique that incorporates a priori information provided by the GLM, and produces eigencomponents that minimise the within group variation while maximising the between group variation. It is performed on a multidimensional contrast (e.g., like that used to specify F-test effects of interest), comprising parameter estimates for the five experimental conditions as well as subject specific columns to remove the between subject variability. The resulting eigencomponents are orthogonal to each other, and show the greatest difference between the parameter estimates in a multivariate network. The method is purely descriptive, aimed to explore the relationship between the predictors and the data, and to help determine linear combinations of the predictors that best summarize the correlation structure.

Figure S2

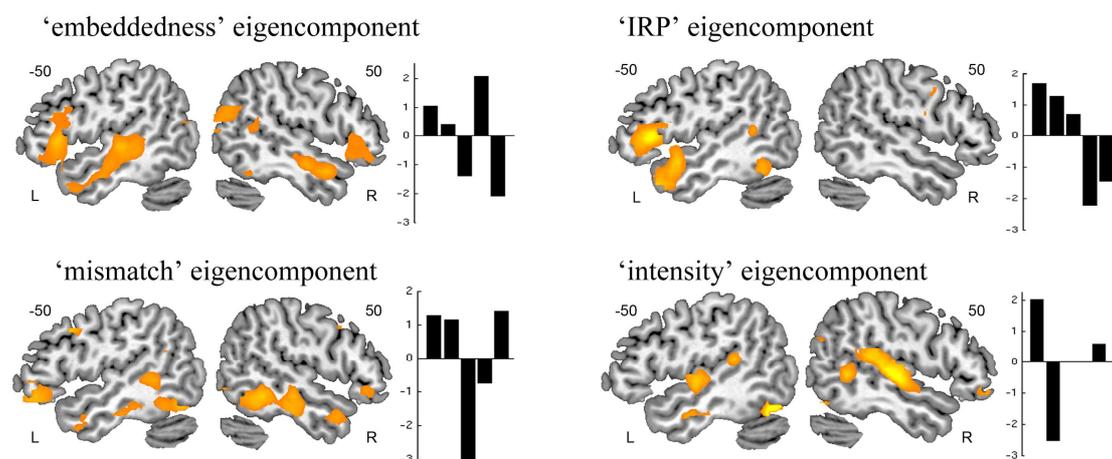


Fig S2. MLM Eigencomponents: Four eigencomponents were obtained in total. In addition to the two described in the manuscript ('embeddedness', which account for 59% of variance; and 'IRP', which accounts for 11%), another component dissociates words in the *blend* condition from the other four conditions, explaining 23% of variance. This is likely to reflect the activation triggered by 'mismatch': the presence of the IRP ending in *blend* triggers combinatorial processing, but the remaining *blen* is not a legitimate word in English and cannot be mapped onto stored lexical representation. The failure of this attempted mapping may trigger the strongly bilateral fronto-temporal activation associated with this component, similar to the one observed for the 'general processing demands/embeddedness' component.

The fourth component dissociates words from *prayed* and *claim* conditions from the words in the *trade* condition, and explains 8% of variance. Overlaid on the brain, activation is seen primarily in auditory cortices in both hemispheres. This may reflect the marginally significant variation in the acoustic intensity of these words sets, measured in dB [means per condition: *prayed* 59.36; *trade* 56.55; *blend* 57.54; *claim* 59.50; *dream* 59.74; $F(4,195)=2.08$, $p=.084$]. All components are shown thresholded at $t>1$.

Bar graphs show how the original predictors (conditions) weight on the linear combinations expressed in an eigencomponent. They are in the unit of the beta images, and can be compared across conditions within a given component. The order of bars in each graph is *prayed*, *trade*, *blend*, *claim*, *dream* (from left to right).

3. Univariate analyses: effects at lower threshold

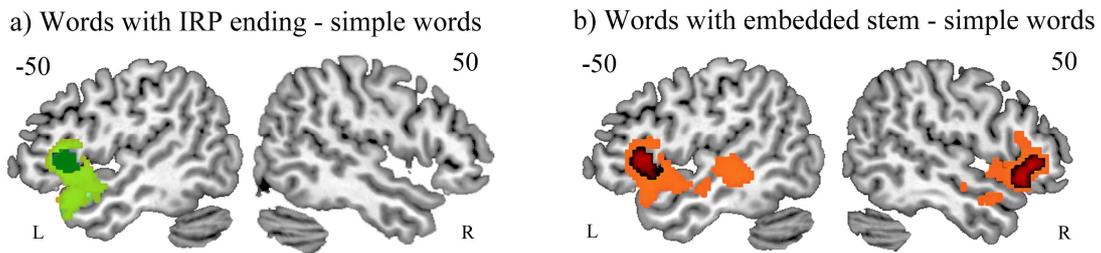


Fig S3. Univariate analyses: Activation at $p < .001$ and $p < .01$ voxel-level thresholds, significant at $p < .05$ cluster-level corrected for multiple comparisons, for (a) words with IRP endings minus simple words and (b) words with embedded stem minus simple words. Lighter colours show the activation at the lower ($p < .01$) threshold.

4. Overlap in LIFG activation

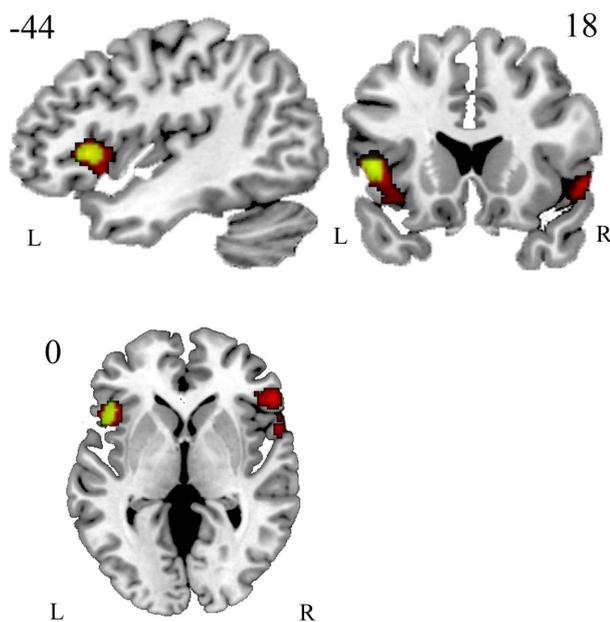


Fig S4. Activation overlap: Overlap between the IRP-related activation and the embedded stem-related activation. IRP-related activation (BA 45, light green/yellow) is fully included within the L embedded stem-related activation (BA 45/47, dark red).

5. Analyses of cohort effects and of non-initial embeddedness

a) Cohort properties of stimuli

In addition to the analyses conducted in terms of onset-embedded competitor stems, we checked the cohort properties of the stimuli, matching these across conditions and testing for effects in the behavioural and neuro-imaging results.

(i) Focusing on the relative weight of cohort competitors as estimated by their frequency, we used the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995) to obtain a measure of the summed frequency of all the cohort members for each test items (cohort members were defined as words (of whatever length) sharing the initial CV/CCV of the word in question). An ANOVA showed that there was no difference in summed cohort frequencies across the five conditions [$F(4,195) < 1$].

(ii) As a more specific index of cohort competition for single items, we looked at the cohort frequency ratio, defined as the ratio of frequency of the specific word to the summed cohort frequency for that word. Here again there was no difference across conditions [$F(4,195) = 1.5$, $p > .1$].

(iii) In addition, we correlated the summed cohort frequency and the cohort ratio with the behavioural scores (reaction times) from the scanner. The results showed no significant correlation.

(iv) Finally, we entered the cohort ratio, on an item-by-item basis, as a parametric modulator in the imaging analysis. There were no regions that showed significant modulation by this variable.

b) Non-initial embeddedness

Non-initial embeddedness is a common occurrence in language, and this is reflected in the stimuli: several words in each condition contained mid or late (rhyme) embedded words. These were often frequent short words such as *end*, *if*, *in*, *aim*, etc, which appear in words from all five conditions, often more than one (e.g. *lay*, *lame*, *aim* in *claim*). We tested whether the numbers and the frequency of these words affect the processing of the word they are embedded in.

(i) Using the CELEX database, we extracted all non-initial embedded words and their frequencies for each of the test items. We then correlated the behavioral scores (reaction times) from the scanner with the number of non-

initial embedded words for each item, as well as with the ratio of the frequency of the specific word to the summed frequency of all its late embedded words (labeled the non-initial embeddedness ratio). There were no significant correlations.

(ii) We tested the distribution of the non-initial embeddedness ratio across the five conditions. An ANOVA showed no difference across conditions [$F(4,195) < 1$].

(iii) We also entered the non-initial embeddedness ratio as a parametric modulator in the imaging analysis, on an item-by-item basis. No regions showed significant modulation by this variable.

Table S1

a) Regions supporting complex acoustic processing; b) regions showing hemispheric laterality effects. Throughout, results were thresholded at $p < .001$ voxel level and clusters that survived $p < .05$ corrected for multiple comparisons were considered significant. MNI coordinates are reported. The highest three peaks within a cluster are shown, with the most significant marked in bold. MuR: Musical Rain.

<i>Regions</i>	<i>Cluster</i>		<i>Voxel</i>	<i>MNI Coordinates</i>		
	<i>P_{correct}</i>	<i>Extent</i>	<i>Z</i>	<i>x</i>	<i>y</i>	<i>z</i>
<i>(a) MuR - null</i>						
L Superior temporal gyrus (BA 42)	0.000	1575	4.49	-66	-30	14
L Superior temporal gyrus (BA 22)			4.19	-44	-40	18
L Heschl's gyrus (BA 41)			4.10	-44	-28	16
R Superior temporal gyrus (BA 42)	0.000	1190	3.98	68	-34	14
R Superior temporal gyrus (BA 22)			3.95	62	-16	4
R Superior temporal gyrus (BA 22)			3.80	58	-36	12
<i>(b) Laterality</i>						
L Insula	0.000	819	4.69	-40	0	12
L Heschl's gyrus (BA 41)			4.56	-38	-24	10
L Heschl's gyrus (BA 41)			4.29	-38	-34	14

Table S2

a) Regions supporting speech-driven lexical processing; b) hemispheric laterality of this process.

<i>Regions</i>	<i>Cluster</i>		<i>Voxel</i>	<i>MNI Coordinates</i>		
	<i>P_{correct}</i>	<i>Extent</i>	<i>Z</i>	<i>x</i>	<i>y</i>	<i>z</i>
<u>(a) Words - MuR</u>						
L Middle temporal gyrus (BA 21)	0.000	1710	5.41	-64	-14	-6
L Middle temporal gyrus (BA 21)			4.58	-66	-26	0
L Anterior temporal cortex (BA 38)			4.48	-44	16	-32
R Middle temporal gyrus (BA 21)	0.000	536	4.38	60	10	-16
R Middle temporal gyrus (BA 21)			4.26	64	2	-16
R Middle temporal gyrus (BA 21)			3.90	68	-10	-4
L Fusiform gyrus (BA 20)	0.000	477	4.65	-38	-22	-18
L Fusiform gyrus (BA 37)			4.47	-32	-36	-20
L Hippocampus			4.42	-40	-34	-4
L Angular gyrus (BA 39)	0.000	362	4.05	-42	-70	26
L Angular gyrus (BA 39)			3.95	-36	-64	26
L Middle occipital gyrus (BA 19)			3.66	-40	-80	32
R Hippocampus	0.000	264	5.11	34	-34	-2
R Hippocampus			4.48	38	-30	-8
R Fusiform gyrus (BA 37)			4.24	36	-40	-8
R Angular gyrus (BA 39)	0.002	209	3.85	54	-68	28
R Angular gyrus (BA 39)			3.82	52	-74	16

R Angular gyrus (BA 39)			3.51	48	-74	32
R Anterior cingulate (BA 32)	0.005	168	4.30	8	22	-10
L Anterior cingulate (BA 32)			4.22	-8	26	-10
R Anterior cingulate (BA 32)			3.66	12	30	-8
<i>(b) Laterality</i>						
L Precentral gyrus (BA 6)	0.000	127	4.33	-40	8	40
L Pars opercularis (BA 44)			4.02	-52	10	34
L angular gyrus (BA 39)	0.000	110	4.51	-40	-74	26
L Pars orbitalis (BA 47)	0.000	104	3.93	-42	28	-14
L Pars orbitalis (BA 47)			3.77	-50	24	-10
L Pars orbitalis (BA 47)			3.51	-36	30	-8
L Pars Triangularis (BA 45)	0.021	48	3.88	-54	24	4

Table S3

Regions showing processing differences as a function of type of lexical complexity: a) Simple words; b) Words with a potential inflectional ending (Inflectional Rhyme Pattern); c) Words with embedded stems. To show the full extent of the IRP-related and stem-related activation, in b) and c) we also report the subsequent most significant peaks from the lower threshold of $p < .01$. These peaks are marked by an asterisk.

<i>Regions</i>	<i>Cluster</i>		<i>Voxel</i>	<i>MNI Coordinates</i>		
	<i>P_{correct}</i>	<i>Extent</i>	<i>Z</i>	<i>x</i>	<i>y</i>	<i>z</i>
<i>(a) Simple -MuR</i>						
L Middle temporal gyrus (BA 21)	0.000	486	5.02	-60	-16	-6
L Middle temporal gyrus (BA 21)			4.29	-64	-6	-4
L Superior temporal gyrus (BA 22)			4.19	-66	-22	-2
R Medial temporal gyrus (BA 21)	0.009	163	4.92	38	-40	-8
R Parahippocampus (BA 37)			4.29	40	-32	-4
<i>(b) IRP present - simple</i>						
LIFC Pars triangularis (BA 45)	0.015	111	3.82	-48	18	4
L superior temporal pole (BA 38)*			2.94	-58	12	-10
L Superior temporal gyrus (BA22)*			2.85	-46	6	-22
<i>(c) Embedded -simple</i>						
RIFC Pars orbitalis (BA 47)	0.000	321	4.50	54	30	0
RIFC Pars orbitalis (BA 47)			4.19	54	22	-8
R superior temporal pole (BA 38)			3.85	56	10	-4
RIFC Pars orbitalis (BA 47) *			3.79	46	28	-4
R Superior temporal gyrus (BA22)*			3.53	56	-2	-10

R Middle temporal gyrus (BA 21)*			3.52	66	-18	-6
LIFC Pars orbitalis (BA 47)	0.000	247	4.43	-44	22	2
LIFC Pars orbitalis (BA 47)			3.31	-36	18	-10
L Middle temporal gyrus (BA 21)*			3.79	-48	-34	-6
L Superior temporal gyrus (BA22)*			3.56	-66	-40	6
L Middle temporal gyrus (BA 21)*			3.39	-60	-24	-4

Table S4

Regions showing modulation by lexical competition.

<i>Regions</i>	<i>Cluster</i>		<i>Voxel</i>	<i>MNI Coordinates</i>		
	<i>P_{correct}</i>	<i>Extent</i>	<i>Z</i>	<i>x</i>	<i>y</i>	<i>z</i>
LIFC Pars orbitalis (BA 47)	0.013	143	4.02	-40	14	-2
L Insula			3.19	-38	6	0
R Insula	0.065*	87	3.98	46	6	-4
RIFC Pars orbitalis (BA 47)			3.61	42	16	-8

* p=.05 at voxel level threshold of p<.005

10. Experimental stimuli for Conditions 1-5 (British English)

Cond.1	Cond.2	Cond.3	Cond.4	Cond.5
bashed	beard	bind	barge	brisk
blamed	brand	bird	binge	bulb
blared	card	bleed	claim	bulge
blurred	chest	blend	clasp	cage
blushed	cleft	blind	coin	chain
boiled	crest	bond	crone	clean
bowed	crude	breed	cusps	cream
cared	deft	bride	drive	cringe
chewed	fade	broad	film	crown
creaked	fast	cloud	frame	deem
cried	feed	code	gain	desk
dared	fend	creed	game	dream
fried	feud	crust	grain	dusk
glared	fold	deed	grange	fringe
gleamed	fund	fiend	grasp	gown
hurled	gold	fist	grove	green
jeered	grade	fraud	gunge	harm
joined	graft	frost	helm	hinge
kicked	grand	ghost	keen	husk
laid	guide	grind	kiln	learn
mashed	hide	guard	lawn	lounge
peered	jade	hard	mosque	lunge
prayed	mild	hound	nine	nerve
purred	mould	lard	palm	plunge
roared	nude	lend	phone	prove
rolled	proud	loud	prime	prune
sailed	raid	nest	rage	queen
saved	rend	pond	range	rasp
shared	ride	rind	rhyme	realm
showed	seed	shade	serve	salve
shunned	shrewd	shield	shelve	shame
sniffed	slide	shift	shine	shave
stared	tend	shroud	sign	sleeve
stayed	trade	swede	singe	task
stirred	tuft	third	spine	term
thawed	tweed	vest	sponge	vague
walked	wand	void	swarm	valve
warned	ward	wound	tinge	wasp

washed
yawned

weird
wind

yard
yield

tone
twinge

wisp
zone