



Published in final edited form as:

Neuroimage. 2020 October 01; 219: 116925. doi:10.1016/j.neuroimage.2020.116925.

No evidence for differences among language regions in their temporal receptive windows

Idan A. Blank^{*,1},

Evelina Fedorenko

Department of Brain and Cognitive Sciences and McGovern Institute for Brain Research,
Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

Abstract

The “core language network” consists of left frontal and temporal regions that are selectively engaged in linguistic processing. Whereas functional differences among these regions have long been debated, many accounts propose distinctions in terms of representational grain-size—e.g., words *vs.* phrases/sentences—or processing time-scale, i.e., operating on local linguistic features *vs.* larger spans of input. Indeed, the topography of language regions appears to overlap with a cortical hierarchy reported by Lerner et al. (2011) wherein mid-posterior temporal regions are sensitive to low-level features of speech, surrounding areas—to word-level information, and inferior frontal areas—to sentence-level information and beyond. However, the correspondence between the language network and this hierarchy of “temporal receptive windows” (TRWs) is difficult to establish because the precise anatomical locations of language regions vary across individuals. To directly test this correspondence, we first identified language regions in each participant with a well-validated task-based localizer, which confers high functional resolution to the study of TRWs (traditionally based on stereotactic coordinates); then, we characterized regional TRWs with the naturalistic story listening paradigm of Lerner et al. (2011), which augments task-based characterizations of the language network by more closely resembling comprehension “in the wild”. We find no region-by-TRW interactions across temporal and inferior frontal regions, which are all sensitive to both word-level and sentence-level information. Therefore, the language network as a whole constitutes a unique stage of information integration within a broader cortical hierarchy.

1. Introduction

Language comprehension engages a cortical network of frontal and temporal brain regions, primarily in the left hemisphere (Binder et al., 1997; Bates et al., 2003; Fedorenko et al., 2010; Menenti et al., 2011). There is ample evidence that this “core language network” is language-selective and is not recruited by other mental processes (Fedorenko and Varley,

^{*}Corresponding author. iblack@psych.ucla.edu (I.A. Blank).

¹Present address: Department of Psychology, University of California Los-Angeles, Los Angeles, CA, 90095, USA.

CRediT authorship contribution statement

Idan A. Blank: Conceptualization, Methodology, Software, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Evelina Fedorenko:** Conceptualization, Methodology, Investigation, Data curation, Writing - review & editing, Supervision, Project administration, Funding acquisition.

2016; see also Pritchett et al., 2018; Ivanova et al., 2019; Jouravlev et al., 2019), indicating that it either employs cognitively unique representational formats and/or implements algorithms distinct from those recruited in other cognitive domains. Nonetheless, the functional architecture of this network—i.e., the division of linguistic labor among its constituent regions—remains highly debated. On the one hand, some neuroimaging studies have suggested that different linguistic processes are localized to distinct, and sometimes focal, subsets of this network (e.g., Stowe et al., 1998; Vandenberghe et al., 2002; Bornkessel et al., 2005; Humphries et al., 2006; Caplan et al., 2008; Snijders et al., 2009; Meltzer et al., 2010; Pallier et al., 2011; Brennan et al., 2012; Goucha and Friederici, 2015; Zhang and Pykkänen, 2015; Kandylaki et al., 2016; Frank and Willems, 2017; Wilson et al., 2018; Bhattasali et al., 2019). Other studies have, on the other hand, reported that different linguistic process (e.g., both lexical and combinatorial) are widely distributed across the network and are spatially overlapping (e.g., Keller et al., 2001; Vigneau et al., 2006; Fedorenko et al., 2012b; Bautista and Wilson, 2016; Blank et al., 2016; Fedorenko et al., 2020; Siegelman et al., 2019). Similar conundrums regarding the mapping of linguistic representations and processes onto distinct vs. shared circuits characterize the neuropsychological (patient) literature (e.g., Caplan et al., 1996; Dick et al., 2001; Bates et al., 2003; Wilson and Saygın, 2004; Grodzinsky and Santi, 2008; Tyler et al., 2011; Duffau et al., 2014; Mesulam et al., 2015; Mirman et al., 2015; Fridriksson et al., 2018; Matchin and Hickok, 2019).

Proposals for the functional architecture of the core language network vary substantially from one another in the theoretical constructs posited and the mapping of those constructs onto brain regions (for examples, see Friederici, 2002; Hickok and Poeppel, 2004; Ullman, 2004; Grodzinsky and Friederici, 2006; Hickok and Poeppel, 2007; Bornkessel-Schlesewsky and Schlesewsky, 2009; Friederici, 2011, 2012; Poeppel et al., 2012; Price, 2012; Bornkessel-Schlesewsky and Schlesewsky, 2013; Hagoort, 2013). Such differences notwithstanding, the majority of accounts share a common, fundamental hypothesis: different language regions integrate incoming linguistic input at distinct timescales. This hypothesis may take different forms: in some accounts, the processing of linguistic representations of different grain size (e.g., phonemes, morphemes/syllables, words, phrases/clauses, and sentences) is respectively mapped onto distinct regions; in other accounts, some region(s) function as mental lexicons (“memory”) that store smaller combinable linguistic units, whereas other regions combine these units into larger structural and meaning representations (“online processing”/“unification”/“composition”). Yet all forms of this hypothesis, while varying considerably in critical details, make the same general prediction: that a functional dissociation among language regions would manifest as differences in their respective timescales for processing and integration.

A brain region’s integration timescale constrains the amount of preceding context that influences the processing of current input. A relatively short integration timescale entails that the incoming signal is integrated with its local context, with more global context exerting little or no influence, whereas a longer integration timescale entails sensitivity to broader contexts extending farther into the past. The amount of context that a brain region is sensitive to governs how closely that region “tracks” input that deviates from well-formedness (e.g., Hasson et al., 2008). For example, a brain region with a short

integration timescale (e.g., on the order of syllables or morphemes) should reliably track any locally well-formed input even in the face of coarser, global disorder (morphemes/syllables can be extracted even from ungrammatical sequences of unrelated words); but a region with a longer integration timescale (e.g., on the order of phrases or clauses) could not reliably track such locally intact but globally incoherent input (phrases/clauses would be difficult or impossible to identify in such sequences). Therefore, a straightforward prediction that follows from the general “different processing timescales” hypothesis is that different language regions should exhibit distinct patterns of tracking for input scrambled at different grain levels (i.e., coarser, more global disruptions that preserve local information vs. finer, more local violations, as described in the examples above).

Indeed, such a pattern of regional response profiles consistent with distinct integration timescales has been reported in a set of left temporal and frontal areas, whose topography appears to overlap with the core language network (Lerner et al., 2011). Specifically, Lerner and colleagues presented participants with a naturalistic spoken story (“intact story”) along with several, increasingly scrambled versions of it: a list of unordered paragraphs (“paragraph list”), a list of unordered sentences (“sentence list”), a list of unordered words (“word list”), and the audio recording played in reverse (“reverse audio”). As participants listened to each of these stimuli, fluctuations in the fMRI BOLD signal were recorded, and the reliability of voxel-wise input tracking was then evaluated. Following (Hasson et al., 2004, 2008), Lerner and colleagues reasoned that if neurons in a given voxel could reliably track a certain stimulus, then the resulting signal fluctuations would be stimulus-locked and, thus, similar across individuals; in contrast, untrackable input would elicit fluctuations that would not be reliably related to the stimulus and, thus, would differ across individuals. Therefore, the authors computed for each voxel and stimulus an inter-subject correlation (ISC; Hasson et al., 2004) of BOLD signal fluctuations. Their novel approach revealed a hierarchy of integration timescales (or “temporal receptive windows”; TRWs) extending from mid-temporal regions both anteriorly and posteriorly along the temporal lobe and on to frontal regions.

Mid-temporal regions early in the hierarchy reliably tracked all stimuli including the reverse audio and the word list conditions, which indicated a very short TRW (~phoneme or below). A little more posteriorly and anteriorly, temporal regions tracked all stimuli except for the reverse audio, indicative of sensitivity to sub-word (e.g., morpheme/syllable) or word-level information. Further posterior and anterior temporal regions could only track lists of sentences or paragraphs (but not word lists), indicative of sensitivity to phrase/clause- or sentence-level information. And, finally, some frontal regions exhibited this same pattern of sensitivity to phrase/clause/sentence information, with yet others reliably tracking only paragraph (but not sentence) lists, indicative of sensitivity to information above the sentence level (a very long TRW).

A hierarchy of integration timescales is an appealing organizing principle of the language network (DeWitt and Rauschecker, 2012; Bornkessel-Schlesewsky et al., 2015; Hasson et al., 2015; Chen et al., 2016; Baldassano et al., 2017; Yeshurun et al., 2017a; Sheng et al., 2018). Nevertheless, there are several reasons to question the putative correspondence between this hierarchy and the set of language-selective cortical regions. The first issue

is neurobiological: the process of TRW characterization described above is carried out on a voxel-by-voxel basis and, hence, crucially relies on the assumption that a given voxel houses the same functional circuits across individuals, but this assumption is demonstrably invalid. Significant inter-individual variability characterizes the mapping of function onto macro-anatomy (Duffau, 2017; Vázquez-Rodríguez et al., 2019; Frost and Goebel, 2012; Tahmasebi et al., 2012), and this variability is especially problematic when functionally distinct regions lie in close proximity to one another, as is the case in both the temporal and frontal lobes (Jones and Powell, 1970; Gloor, 1997; Wise et al., 2001; Chein et al., 2002; Fedorenko et al., 2012a; Deen et al., 2015; Braga et al., 2019; for a related review, see: Fedorenko and Blank, 2020). In these areas, the same stereotactic coordinate may be part of the core language network in one brain but part of a functionally distinct network in another, which severely complicates the interpretation of voxel-based inter-subject correlations in BOLD signal fluctuations as markers of input tracking by a specific functional circuit.

The second issue is statistical. Even if different language regions showed evidence of differing integration timescales at the descriptive level, direct statistical comparisons across their response profiles would be required in order to establish that they are indeed functionally distinct. For instance, when the tracking of the word list condition is significant in one region but not in another, the difference between these two regions might itself still be non-significant (Nieuwenhuis et al., 2011). In other words, a region-by-condition interaction test is a crucial piece of statistical evidence in support of different integration timescales among the regions of the core language network, but such a test has hitherto been missing.

The third issue pertains to psycholinguistic theory and data. Although there is little doubt that comprehension proceeds via a cascaded integration of input along increasingly longer timescales (constructing larger meaningful units out of smaller ones; see, e.g., Christiansen and Chater, 2016), different stages of this process need not rely on qualitatively distinct mental structures or memory stores. Instead, language processing appears to operate over a continuum of merely “quantitatively” different representations that straddle the traditionally postulated boundaries between sounds and words (Farmer et al., 2006; Bradlow and Bent, 2008; Maye et al., 2008; Trude and Brown-Schmidt, 2012; Schmidtke et al., 2014) and between words and larger constructions and combinatorial rules (Clifton et al., 1984; MacDonald et al., 1994; Trueswell et al., 1994; Garnsey et al., 1997; Traxler et al., 2002; Reali and Christiansen, 2007; Gennari and MacDonald, 2008) (see also Joshi et al., 1975; Schabes et al., 1988; Goldberg, 1995; Bybee, 1998; Jackendoff, 2002; Culicover and Jackendoff, 2005; Wray, 2005; Bybee, 2010; Snider and Arnon, 2012; Jackendoff, 2007; Langacker, 2008; Christiansen and Arnon, 2017). By extension from cognition to its neural implementation, linguistic representations of different grain sizes, and their processing, need not be spatially segregated in the cortex across distinct regions.

Therefore, the current study directly tested for a functional dissociation among core language regions in terms of their temporal receptive windows. To this end, and to address the methodological issues discussed above, we synergistically combined two neuroimaging paradigms with complementary strengths: a traditional, task-based design and a naturalistic, task-free design. First, we used a well-validated localizer task to identify regions of the core language network individually in each participant (Fedorenko et al., 2010). This approach

allowed us to establish correspondence across brains based on functional response profiles (Saxe et al., 2006) rather than stereotaxic coordinates in a common space, thus augmenting the common voxel-based methodologies for studying temporal receptive windows (Hasson et al., 2008). Then, we characterized the temporal receptive window of each language region by using the naturalistic story and its scrambled versions from Lerner et al. (2011). This paradigm broadly samples the space of representations and computations engaged during comprehension and, thus, tests the “different processing timescales” hypothesis in its most general formulation, decoupled from more detailed theoretical commitments. It therefore augments task-based paradigms, which rely on materials and tasks that isolate particular mental processes tied to specific theoretical constructs. Finally, we computed inter-subject correlations for each condition in each functionally localized language region, and tested for a region-by-condition interaction to directly compare the resulting temporal receptive window profiles across the network. In sum, our combined approach both (i) enjoys the increased ecological validity of naturalistic, “task-free” neuroimaging paradigms that mimic comprehension “in the wild” (Maguire, 2012; Sonkusare et al., 2019); and (ii) ensures the functional interpretability of the studied regions by harnessing a participant-specific localizer task instead of relying on precarious “reverse inference” from anatomy back to function (Poldrack, 2006, 2011).

2. Materials and methods

2.1. Participants

Twenty participants (12 females) between the ages of 18 and 47 (median = 22), recruited from the MIT student body and the surrounding community, were paid for participation. All participants were native English speakers, had normal hearing, and gave informed consent in accordance with the requirements of MIT’s Committee on the Use of Humans as Experimental Subjects (COUHES). One participant was excluded from analysis due to poor behavioral performance on a postscan assessment and neuroimaging data quality; the results below include the remaining 19 participants.

All participants but one had a left-lateralized language network, as determined based on visual inspection of their language localizer data (see section 2.2.1). The remaining participant was left-handed (Willems et al., 2014) and had a right-lateralized network; therefore, for this participant only, language fROIs were defined in the right hemisphere.

2.2. Design, materials and procedure

Each participant performed the language localizer task (Fedorenko et al., 2010) and, for the critical experiment, listened to all five versions of a narrated story (cf. Lerner et al., 2011, where different subsets of the sample listened to different subsets of the stimulus set). The localizer and critical experiment were run either in the same scanning session (13 participants) or in two separate sessions (6 participants, who have previously performed the localizer task while participating in other studies) (see Mahowald and Fedorenko, 2016 for evidence of high stability of language localizer activations over time; see also Braga et al., 2019). In each session, participants performed a few other, unrelated tasks, with scanning sessions lasting 90–120min.

2.2.1. Language localizer task—Regions in the core language network were localized using a passive reading task that contrasted sentences (e.g., DIANE VISITED HER MOTHER IN EUROPE BUT COULD NOT STAY FOR LONG) and lists of unconnected, pronounceable nonwords (e.g., LAS TUPING CUSARISTS FICK PRELL PRONT CRE POME VILLPA OLP WORNIST CHO) (Fedorenko et al., 2010). Each stimulus consisted of 12 words/nonwords, presented at the center of the screen one word/nonword at a time at a rate of 450ms per word/nonword. Each trial began with 100ms of fixation and ended with an icon instructing participants to press a button, presented for 400ms and followed by 100ms of fixation, for a total trial duration of 6s. The button-press task was included to help participants remain alert and focused throughout the run. Trials were presented in a standard blocked design with a counterbalanced order across two runs. Each block, consisting of 3 trials, lasted 18s. Fixation blocks were evenly distributed throughout the run and lasted 14s. Each run consisted of 8 blocks per condition and 5 fixation blocks, lasting a total of 358s. (A version of this localizer is available for download from <https://evlab.mit.edu/funcloc/download-paradigms>.)

The *sentences > nonwords* contrast targets high-level aspects of language, to the exclusion of perceptual (speech/reading) and motor-articulatory processes (for a discussion, see Fedorenko and Thompson-Schill, 2014; Fedorenko, in press). We chose to use this particular localizer contrast for compatibility with other past and ongoing experiments in the Fedorenko lab and other labs using similar localizer contrasts. For the current study, the main requirements from the localizer contrast were that it neither be under-inclusive (i.e., fail to identify some regions of the core language network) nor over-inclusive (i.e., identify some regions that lie outside of, and are functionally distinct from, the core language network). Below, we address each requirement in turn.

First, to avoid under-inclusiveness, the contrast should identify regions engaged in a variety of high-level linguistic processes, from ones that might depend on relatively local information integration (e.g., single-word processing) to those that might depend on more global integration (e.g., processing of multi-word constructions, online composition). Because sentences differ from nonword lists in requiring processing at both the word level and phrase/clause/sentence level, the contrast's content validity is appropriate in terms of capturing linguistic processes across multiple scales. Moreover, this localizer contrast has been extensively validated over the past decade, and shown to identify regions that all exhibit sensitivity to word-, phrase/clause-, and sentence-level semantic and syntactic processing (Fedorenko et al., 2010, 2012b, 2020; Blank et al., 2016; Mollica et al., 2018), which are the levels we focus on as described in the Results section. For instance, the regions identified with this localizer all exhibit reliable effects for narrower contrasts, like sentences *vs.* word lists; sentences *vs.* “Jabberwocky” sentences (where content words have been replaced with nonwords); word lists *vs.* nonword lists; and “Jabberwocky” sentences *vs.* nonword lists (similar patterns obtain in electrocorticographic data with high temporal resolution: Fedorenko et al., 2016). In addition, contrasts that are broader than *sentences > nonwords* and that do not subtract out phonology and/or discourse-level processes (e.g., a contrast between natural spoken paragraphs and their acoustically degraded versions: Scott et al., 2016; Ayyash et al., in prep.) identify the same network. Moreover, activations to

the *sentences* > *nonwords* contrast exhibit extremely tight overlap with a fronto-temporal network identified solely based on resting-state data (Braga et al., 2019; Branco et al., 2020). Therefore, if we do not observe functional dissociations among language regions in their respective TRWs, it would not be simply because the language localizer subsamples a functionally homogeneous subset of regions out of a larger network.

Second, to avoid over-inclusiveness, the contrast should not identify functional networks that are distinct from the core language network and might be recruited during online comprehension for other reasons (e.g., task demands, attention, episodic encoding, non-verbal knowledge retrieval, or mentalizing). Whereas there are many potential differences between the processing of sentences vs. nonwords that might engage such non-linguistic processes, the identified regions exhibit robust language selectivity in their responses, showing little or no response to non-linguistic tasks (Fedorenko et al., 2011; Fedorenko et al., 2012a; Pritchett et al., 2018; Ivanova et al., 2019; Jouravlev et al., 2019; For a review, see Fedorenko and Varley, 2016; Scott, 2020). Moreover, whereas these regions are synchronized with one another during naturalistic cognition, they are strongly dissociated from other brain networks (Blank et al., 2014; Paunov et al., 2019; for evidence from inter-individual differences, see: Mineroff et al., 2018). Therefore, if we observe functional dissociations among regions in their respective TRWs, it would not be simply because the localizer oversamples a functionally heterogeneous set of regions that extend beyond the core language network. In sum, the contrast we use appears to identify a network that is a “natural kind”.

The evidence reviewed above provides strong support for both convergent and discriminant validity of the language localizer. In addition, this localizer generalizes across materials, tasks (passive reading vs. a memory probe task), and modality of presentation (visual and auditory: Fedorenko et al., 2010; Braze et al., 2011; Vagharchakian et al., 2012; Deniz et al., 2019). Presentation modality is particularly important because the main task of the current study relied on spoken language comprehension.

2.2.2. Critical experiment—Participants listened to the same materials that were originally used to characterize the cortical hierarchy of temporal receptive windows. These materials were based on an audio recording of a narrated story (“Pie-Man”, told by Jim O’Grady at an event of “The Moth” group, NYC). The conditions included (i) the intact audio; (ii) three “scrambled” versions of the story that differed in the temporal scale of incoherence, namely, lists of randomly ordered paragraphs, sentences, or words, respectively; and (iii) a reverse audio version. The last condition served as a low-level control, because reverse speech is acoustically similar to speech and is similarly processed (by lower-level auditory regions; Lerner et al., 2011), but does not carry linguistic information beyond the phonetic level (Kimura and Folb, 1968; Koeda et al., 2006; but see Norman-Haignere et al., 2015). We note that these conditions map only onto vague notions of “words”, “sentences”, and “paragraphs”; their mapping onto psycholinguistic constructs such as “phonemes”, “syllables”, “morphemes”, “phrases”, or “clauses” remains under-determined (for instance, the word-list condition differs from the reverse audio condition in the presence of many phonemes, as well as syllables, morphemes, and word-level lexical entries).

To render these materials suitable for our existing scanning protocol, which used a repetition time of 2s (see section 2.3.1.), the silence and/or music period preceding and following each stimulus were each extended from 15s to 16s so that they fit an integer number of scans. These periods were not included in the analyses reported below. In addition, the two longest paragraphs in the paragraph-list stimulus were each split into two sections, and one section was randomly repositioned in the stream of shuffled paragraphs. No other edits were made to the original materials.

Participants listened to the materials (one stimulus per run) over scanner-safe headphones (Sensimetrics, Malden, MA), in one of two orders: for 10 participants, the intact story was played first and was followed by increasingly finer levels of scrambling (from paragraphs to sentences to words). For the remaining 9 participants, the word-list stimulus was played first and was followed by decreasing levels of scrambling (from sentences to paragraphs to the intact story). The reverse story was positioned either in the middle of the scanning session or at the end, except for one participant for whom we could not fit this condition in the scanning session.

At the end of the scanning session, participants answered 8 multiple-choice questions concerning characters, places, and events from particular points in the narrative, with foils describing information presented elsewhere in the story. All participants demonstrated good comprehension of the story (17 of them answered all questions correctly, and the remaining two had only one error; the 20th participant, excluded from analysis, had 50% accuracy).

2.3. Data acquisition and preprocessing

2.3.1. Data acquisition—Structural and functional data were collected on a whole-body 3 Tesla Siemens Trio scanner with a 32-channel head coil at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. T1-weighted structural images were collected in 176 axial slices with 1 mm isotropic voxels (repetition time (TR) = 2,530ms; echo time (TE) = 3.48ms). Functional, blood oxygenation level-dependent (BOLD) data were acquired using an EPI sequence with a 90° flip angle and using GRAPPA with an acceleration factor of 2; the following parameters were used: thirty-one 4.4 mm thick near-axial slices acquired in an interleaved order (with 10% distance factor), with an in-plane resolution of 2.1 mm × 2.1 mm, FoV in the phase encoding (A >> P) direction 200 mm and matrix size 96 mm × 96 mm, TR = 2,000ms and TE = 30ms. The first 10s of each run were excluded to allow for steady state magnetization.

2.3.2. Data preprocessing—Spatial preprocessing was performed using SPM5 and custom MATLAB scripts. (Note that SPM was only used for preprocessing and basic first-level modeling, aspects that have not changed much in later versions; we used an older version of SPM because data for this study are used across other projects spanning many years and hundreds of participants, and we wanted to keep the SPM version the same across all the participants.) Anatomical data were normalized into a common space (Montreal Neurological Institute; MNI) template, resampled into 2 mm isotropic voxels, and segmented into probabilistic maps of the gray matter, white matter (WM) and cerebrospinal fluid (CSF). Functional data were motion corrected, resampled into 2 mm isotropic voxels,

and high-pass filtered at 200s. Data from the localizer runs were additionally smoothed with a 4 mm FWHM Gaussian filter, but data from the critical experiment runs were not, in order to avoid blurring together the functional profiles of nearby regions with distinct TRWs (we obtained the same results following spatial smoothing in a supplementary analysis).

Additional temporal preprocessing of data from the critical experiment runs was performed using the SPM-based CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012) with default parameters, unless specified otherwise. Five temporal principal components of the BOLD signal time-courses extracted from the WM were regressed out of each voxel's time-course; signal originating in the CSF was similarly regressed out. Six principal components of the six motion parameters estimated during offline motion correction were also regressed out, as well as their first time derivative. Next, the residual signal was bandpass filtered (0.008–0.09 Hz) to preserve relatively low-frequency signal fluctuations, because higher frequencies might be contaminated by fluctuations originating from non-neural sources (Cordes et al., 2001).

We note that bandpass filtering was not used by Lerner et al. (2011). In another supplementary analysis, without filtering, we obtained the same pattern of results reported below. However, the unfiltered time-courses exhibited overall lower reliability across participants. We therefore chose to report the analyses of the filtered data.

2.4. Data analysis

All analyses were performed in MATLAB (The MathWorks, Natick, MA) unless specified otherwise.

2.4.1. Functionally defining language regions in individual participants—Data from the language localizer task were analyzed using a General Linear Model that estimated the voxel-wise effect size of each condition (sentences, nonwords) in each run of the task (the two runs were included in the same GLM model, but all regressors were defined perrun, i.e., the design matrix was block-diagonal). These effects were each modeled with a boxcar function (representing entire blocks) convolved with the canonical Hemodynamic Response Function (HRF). The model also included first-order temporal derivatives of these effects, as well as nuisance regressors representing entire experimental runs and offline-estimated motion parameters. The obtained beta weights were then used to compute the voxel-wise *sentences* > *nonwords* contrast, and these contrasts were converted to *t*-values. The resulting *t*-maps were restricted to include only gray matter voxels, excluding voxels that were more likely to belong to either the WM or the CSF based on the probabilistic segmentation of the participant's structural data.

Functional regions of interest (fROIs) in the language network were then defined using group-constrained, participant-specific localization (Fedorenko et al., 2010). For each participant, the *t*-map of the *sentences* > *nonwords* contrast (pooled across the two runs) was intersected with binary masks that constrained the participant-specific language regions to fall within areas where activations for this contrast are relatively likely across the population. These five masks, covering areas of the left temporal and frontal lobes, were derived from a group-level probabilistic representation of the localizer contrast in an independent set of 220

participants (available for download from: <https://evlab.mit.edu/funcloc/download-parcels>). In order to increase functional resolution in the temporal cortex, where a gradient of multiple distinct TRWs was originally reported by Lerner et al. (2011), the two temporal masks were each further divided in two, approximately along the posterior-anterior axis. The border locations marking these divisions were determined based on an earlier version of the group-level representation of the localizer contrast, obtained from a smaller sample (Fedorenko et al., 2010). In total, seven masks were used (Fig. 1), in the posterior, mid-posterior, mid-anterior, and anterior temporal cortex, the inferior frontal gyrus, its orbital part, and the middle frontal gyrus. (Unlike previous reports from our group that had used an additional mask in the angular gyrus, we decided to exclude this region going forward because it does not appear to be a part of the core language network in either its task-based responses or its signal fluctuations during naturalistic cognition. See, e.g., Blank et al., 2014; Blank et al., 2016; Pritchett et al., 2018; Ivanova et al., 2019; Jouravlev et al., 2019; Paunov et al., 2019).

In each of these masks, a participant-specific fROI was defined as the top 10% of voxels with the highest t -values for the *sentences* > *nonwords* contrast. fROIs within the smallest mask counted 37 voxels, and those within the largest mask—237 voxels (Fig. 1). This top $n\%$ approach ensures that fROIs can be defined in every participant and that their sizes are the same across participants, allowing for generalizable results (Nieto-Castañón and Fedorenko, 2012). In line with much prior work, these language fROIs showed highly replicable *sentences* > *nonwords* effects, estimated using independent portions of the data for fROI definition and response estimation (for all regions, $t_{(18)} > 5.97$, $p < 10^{-4}$, corrected for multiple comparisons using false-discovery rate (FDR) correction (Benjamini and Yekutieli, 2001); Cohen's $d > 1.24$, where this effect size, but not the previous measures, is based on a conservative, independent samples t -test).

We additionally defined a few alternative sets of fROIs for control analyses. First, to ensure that language fROIs were each functionally homogeneous and did not group together sub-regions with distinct TRWs, we also defined alternative, smaller fROIs based on the top 4% of voxels with the highest localizer contrast effects in each mask (these were 15–95 voxels in size). We were also interested in whether TRWs in the core language network differed from those in neighboring regions exhibiting weaker localizer contrast effects. Therefore, we defined fROIs based on the “second-best” 4% of voxels within each mask (i.e., those whose effect sizes were between the 92 and 96 percentile), as well as based on the “third best” (88–92 percentile), “fourth best” (84–88 percentile), and “fifth best” (80–84 percentile) sets.

2.4.2. Main analysis of temporal receptive windows in language fROIs—For each of the five conditions in the critical experiment (intact story, paragraph list, sentence list, word list, and reverse audio), in each of seven fROIs and for each participant, BOLD signal time-series were extracted from each voxel and were then averaged across voxels to obtain a single time-series per fROI, participant, and condition. When extracting these signals we skipped the first 6s (3 vol) following stimulus onset, in order to exclude a potential initial rise of the hemodynamic response relative to fixation; such a rise would be a trivially reliable component of the BOLD signal that might blur differences among conditions and fROIs. In addition, we included 6s of data following stimulus offset, in order to account for the hemodynamic lag (we obtained the same pattern of results in a

supplementary analysis in which we skipped the first 10s and did not include any data post stimulus offset).

To compute ISCs per fROI and condition, we temporally *z*-scored the time-series of all participants but one, averaged them, and computed Pearson's moment correlation coefficient between the resulting group-averaged time-series and the corresponding time-series of the left-out participant. This procedure was iterated over all partitions of the participant pool, producing 19 ISCs per fROI and condition. These ISCs were Fisher-transformed to improve the normality of their distribution (Silver and Dunlap, 1987).

To reiterate the logic detailed in the introduction, the resulting regional ISCs quantify the similarity of regional BOLD signal fluctuations across participants, with high values indicative of regional activity that reliably tracks the incoming input (correlations across participants mirror correlations within a single participant across stimulus presentations; Golland et al., 2007; Hasson et al., 2009; Blank and Fedorenko, 2017). Further, ISCs across the five conditions constitute a functional profile characterizing a region's TRW. Namely, reliable input tracking (i.e., high ISCs) is expected only for stimuli that are well-formed at the timescale over which a given region integrates information; weaker tracking (i.e., low ISCs) is expected for stimuli that are scrambled at that scale and, thus, cannot be reliably integrated.

For descriptive purposes, we first labeled the TRW of each fROI based on the most scrambled condition for which tracking was still statistically indistinguishable from tracking of the intact story. For example, if ISCs in a certain region were uniformly high for the intact story, paragraph list, and sentence list, but were significantly lower for the word list, that region's TRW was labeled as "sentence-level" (because input tracking incurred a cost when well-formedness at that level was violated). Similarly, if ISCs in another region were uniformly high for the intact story, paragraph list, sentence list, and word list, but dropped for the reverse audio, that region's TRW was labeled as "word-level". To thus label TRWs, for each fROI we compared ISCs between the intact story and every other stimulus using dependent-samples *t*-tests ($\alpha = 0.05$, here and in all tests below). The resulting *p*-values were corrected for multiple comparisons using false-discovery rate (FDR) correction (Benjamini and Yekutieli, 2001) across all pairwise comparisons and fROIs.

For our main analysis, we directly compared the pattern of ISCs to the five conditions across the seven language fROIs via a two-way, repeated-measures analysis of variance (ANOVA) with fROI (7 levels) and condition (5 levels) as within-participant factors. The critical test was for a region-by-stimulus interaction. To further interpret our findings, we conducted follow-up analyses as detailed in the Results section. In addition to the parametric ANOVA, we ran empirical permutation tests of reduced residuals (Anderson and Braak, 2003), which are less sensitive to violations of the test's assumptions, and obtained virtually identical results. We chose ANOVA over mixed-effects linear regression because the latter is more conservative due to estimator shrinkage, and we wanted to give any region-by-condition interaction—should one be present—the strongest chance of revealing itself. Nonetheless, inferences remained unchanged when tests were run via linear, mixed-effects regressions

(using the lme4 toolbox in R) with varying intercepts by fROI, condition, and participant (Gelman, 2005).

2.4.3. Controlling for baseline differences in input tracking across fROIs—

When comparing functional responses across brain regions, it is critical to take into account regional differences in baseline responsiveness, because these might mask fROI-by-condition interactions (or explain them away; Nieuwenhuis et al., 2011). For instance, whereas an ANOVA might conclude that a difference between an ISC of 0.5 for the intact story and an ISC of 0.4 for the reverse audio in one region is statistically indistinguishable from a difference between ISCs of 0.2 and 0.1 in another, the former difference constitutes only a 20% decrease whereas the latter constitutes a 50% decrease. Therefore, we corrected for such baseline differences and re-tested for a fROI-by-condition interaction.

To this end, we used regional ISCs for the intact story as a “ceiling” against which to normalize ISCs for the other four conditions: first, all ISCs were converted back to the $[-1,1]$ range using the inverse Fisher transform. Then, we squared all ISCs, thus transforming them from correlations to “percentage of explained variance” (i.e., the fraction of variance in BOLD signal fluctuations from one participant that is explained by the corresponding BOLD signal fluctuations from other participants). Next, we divided squared-ISCs for each of the paragraph list, sentence list, word list, and reverse audio conditions by their corresponding squared-ISCs for the intact story (the division was performed separately for each participant and fROI). This division acted as a “normalization” procedure, where the percentages of explained variance in different conditions were compared against their ceiling value from the intact condition. Finally, we took the square root of the resulting values in order to transform them into (normalized) correlations. Whenever the resulting normalized correlation was greater than 1, it was rounded down to 1. We tested these normalized ISCs for fROI-by-condition interaction, as described in section 2.4.2.

2.4.4. Additional, voxel-based analyses—

Whereas our main analyses examined ISCs in functionally defined, participant-specific fROIs, we also conducted two control analyses for which ISCs were defined using the common, voxel-based approach. Although we believe this approach is disadvantageous and suffers from interpretational limitations (see Introduction), we performed these analyses in order to provide a more comprehensive investigation of TRWs in the core language network.

Our first goal was to replicate the original findings from Lerner et al. (2011) so as to ensure that any differences between our main analyses and this previous study do not result from inconsistencies in the data. For this analysis, following Lerner et al. (2011), we smoothed the (temporally preprocessed) functional scans with a 6 mm FWHM Gaussian kernel. Then, for each of the five conditions in the main experiment, we computed voxel-wise ISCs for the subset of left-hemispheric voxels that met the following three criteria: (i) were more likely to be gray matter than either WM or CSF in at least 2/3 ($n = 13$) of the participants, based on the probabilistic segmentation of their individual, structural data; (ii) were part of the frontal, temporal, or parietal lobes as defined by the AAL2 atlas (Tzourio-Mazoyer et al., 2002; Rolls et al., 2015); and (iii) fell in the cortical mask used for the cortical parcellation in Yeo et al. (2011). We then labeled the TRW of each voxel following the approach of

Lerner et al. (2011), namely, based on the most scrambled stimulus that, across participants, was still tracked significantly above chance (evaluated against a Gaussian fit to an empirical null distribution of ISCs that was generated from surrogate signal time-series; see Theiler et al., 1992). Tests were FDR-corrected for multiple comparisons across conditions and voxels. The resulting map of TRWs was projected onto Freesurfer's average cortical surface in MNI space. No further quantitative tests were performed, as we were only interested in obtaining a visually similar gradient of TRWs to the one previously reported.

Our second goal was an alternative definition of fROIs that relied less heavily on the task-based functional localizer, in order to alleviate any remaining concerns regarding its use (see section 2.2.1.). Here, rather than defining fROIs that maximized the localizer contrast effect and subsequently characterizing their TRWs, we aimed at defining fROIs that directly maximized the ISC profiles consistent with certain TRWs. To avoid circularity from the use of the same data to define fROIs and to estimate their response profiles (Vul and Kanwisher, 2010; Kriegeskorte et al., 2009), we first created two independent sets of ISCs by splitting each BOLD signal time-series and computing ISCs for each half of the data. We then used data from the second half of each stimulus to define fROIs, and data from the first half to compare TRWs across the resulting fROIs. (We used the first half of each stimulus for the critical test because we suspected input tracking would be overall lower in the second half due to participants losing focus, especially for scrambled stimuli without coherent meaning; however, the pattern of results did not depend on which data half was used for fROI definition and which was used for the critical test.)

This analysis proceeded as follows: in each mask (same masks as the ones used in the main analyses), we first labeled the TRW of each voxel as in our main analysis (section 2.4.2.), i.e., based on the most scrambled condition for which tracking was statistically indistinguishable from tracking of the intact story (the alternative labeling scheme described in the second paragraph of this section, based on the most scrambled condition that was still tracked significantly above chance, yielded similar results). We then chose the voxels whose TRW label matched the label of the localizer-based fROI from the main analysis (see the penultimate paragraph of section 2.4.2, and an example below). We sorted these voxels based on the size of the difference between their tracking of the intact story and of the least scrambled condition that was tracked less reliably, and chose the 27 voxels whose *p*-values for that comparison were the smallest (most significant) (27 voxels is the number corresponding to a 3-voxel cubic neighborhood, but we did not constrain these voxels to be contiguous). For example, if the TRW of interest was “sentence-level”, this meant that we focused on voxels (i) whose tracking of the sentence list did not statistically differ from their tracking of the intact story, but (ii) their tracking of the word list was significantly less reliable; a “sentence-level” fROI was then chosen as the 27 voxels showing the most significant differences between ISCs for the intact story and the word list condition (in a dependent-samples *t*-test across participants). Once a fROI was defined this way in each mask, we averaged the ISCs for the first half of each stimulus across its voxels, and compared the resulting ISC profiles across fROIs using two-way, repeated-measures ANOVAs with the factors fROI (7 levels) and condition (5 level), as in our main analyses (section 2.4.2.).

2.4.5. Comparing TRWs between the language network and other functional regions

—To situate data from the core language network in a broader context, we also computed ISCs in other cortical regions. First, we computed ISCs based on BOLD signal time-series averaged across voxels from an anatomically defined mask of lower-level auditory cortex in the anterolateral section of Heschl’s gyrus in the left hemisphere (Tzourio-Mazoyer et al., 2002) (we did not use a functional localizer because lower-level sensory regions show overall better mapping onto macro-anatomy compared to higher-level associative regions: Frost and Goebel, 2012; Tahmasebi et al., 2012; Vázquez-Rodríguez et al., 2019). This region has a short TRW (e.g., Lerner et al., 2011; Honey et al., 2012a), and was therefore expected to track all stimuli equally reliably.

Second, we extracted ISCs from regions of the “episodic” (or “default mode”) network (Gusnard and Raichle, 2001; Raichle et al., 2001; Buckner et al., 2008; Andrews-Hanna et al., 2010; Humphreys et al., 2015), which is engaged in processing episodic information. This network, recruited when we process events as part of a narrative, is expected to integrate input over longer timescales compared to the core language network (Regev et al., 2013; Chen et al., 2016; Margulies et al., 2016; Simony et al., 2016; Yeshurun et al., 2017a, 2017b; Zadbood et al., 2017; Nguyen et al., 2019). To identify fROIs in this network we relied on its profile of deactivation during tasks that tax executive functions for the processing of external stimuli, and used a visuo-spatial working-memory localizer that included a “hard” condition requiring the memorization of 8 locations on a 3×4 grid (Fedorenko et al., 2011, 2013). We defined the contrast *hard* < *fixation* and, following the approach outlined above (section 2.4.1.), chose the top 10% of voxels showing the strongest *t*-values for this contrast in two left-hemispheric masks, located in the posterior cingulate cortex and temporo-parietal junction (these masks were generated based on a group-level probabilistic representation of the localizer task data from 197 participants).

To compare TRWs between the core language network and each of these two other systems, we averaged ISCs for each condition across fROIs in each system and conducted two-way, repeated-measures ANOVAs with system (2 levels: language and auditory/language and episodic) and condition (5 levels) as within-participant factors. The critical test was for a system-by-condition interaction.

3. Results

3.1. No evidence for a region-by-condition interaction among language fROIs in the left inferior frontal and temporal cortices

3.1.1. Main analysis—The main results of the current study are presented in Fig. 2A. We computed inter-subject correlations for each of the five conditions in each of seven core language fROIs, and directly compared the resulting regional profiles of input tracking via a two-way (fROI \times condition), repeated-measures ANOVA. As expected, there was a main effect of condition ($F_{(4,68)} = 32.7$, partial $\eta^2 = 0.66$, $p < 10^{-14}$). Follow-up ANOVAs (FDR-corrected for multiple comparisons) contrasting the intact story to each other condition revealed no overall differences in tracking the intact story and the paragraph list ($F_{(1,18)} = 0$, $\eta^2_p = 0$, $p = 1$) or the sentence list ($F_{(1,18)} = 0.02$, $\eta^2_p = 10^{-3}$, $p = 1$), but weaker tracking of the word list ($F_{(1,18)} = 27.27$, $\eta^2_p = 0.6$, $p < 10^{-3}$) and the reverse audio condition ($F_{(1,18)}$

= 169.90, $\eta^2_p = 0.92$, $p < 10^{-7}$). Furthermore, the sentence list was tracked more reliably than the word list ($F_{(1,18)} = 9.76$, $\eta^2_p = 0.35$, $p = 0.04$) which was, in turn, tracked more reliably than the reverse audio condition ($F_{(1,18)} = 48.56$, $\eta^2_p = 0.74$, $p < 10^{-4}$). In addition, there was a main effect of fROI ($F_{(6,102)} = 19.7$, $\eta^2_p = 0.54$, $p < 10^{-14}$), indicating that some fROIs overall tracked stimuli more strongly than others, a finding we return to in section 3.1.2.

As an initial characterization of the region-wise TRWs, we ran dependent-samples *t*-tests in each fROI to compare ISCs for the intact story and for each of the other conditions. We then identified the most scrambled condition whose tracking was still statistically indistinguishable from tracking of the intact story (correcting for multiple tests across pairwise comparisons and fROIs). These tests indicated that the mid-posterior, mid-anterior, and anterior temporal fROIs each exhibited “word-level” TRWs, with input tracking reliability not incurring a cost when words were randomly ordered, but becoming significantly weaker for the reverse audio condition (for all three regions: $t_{(17)} > 6.54$, Cohen’s $d > 1.58$, $p < 10^{-4}$). The fROIs in the posterior temporal cortex, inferior frontal gyrus, its orbital part, and middle frontal gyrus each exhibited “sentence-level” TRWs, with input tracking reliability not incurring a cost for the sentence list condition, but becoming significantly weaker for the word list condition (for all four regions: $t_{(18)} > 2.8$, $d > 0.66$, $p < 0.04$). Prior to multiple comparison correction, all seven fROIs exhibited “word-level” TRWs. Based on these findings, in several of our analyses below we report tests focusing on the sentence list and word list conditions, which appear to be the locus of potential functional differences across fROIs.

Critically, whereas the ANOVA for a fROI-by-condition interaction in ISCs was significant ($F_{(24,408)} = 1.78$, $\eta^2_p = 0.10$, $p = 0.014$), this interaction was explained by the middle frontal gyrus (MFG) fROI: follow-up analyses testing for an interaction across all regions but one (uncorrected for multiple comparisons, so as to be anti-conservative) failed to reach significance when the MFG was removed ($F_{(20,340)} = 1.32$, $\eta^2_p = 0.07$, $p = 0.16$), but remained significant when each of the other fROIs was removed (for all tests, $F_{(20,340)} > 1.71$, $\eta^2_p > 0.09$, $p < 0.03$). The same results obtained when testing only the sentence list and word list conditions (all seven fROIs: $F_{(6,108)} = 4.32$, $\eta^2_p = 0.19$, $p < 10^{-3}$; MFG fROI excluded: $F_{(5,90)} = 2.24$, partial $\eta^2 = 0.11$, $p = 0.057$; any other fROI excluded: $F_{(5,90)} > 4.11$, $\eta^2_p > 0.18$, $p < 0.002$). Furthermore, the word list condition appeared to drive the interaction across the seven fROIs: when this condition was excluded from analysis, the interaction test was not significant ($F_{(18,306)} = 1.13$, $\eta^2_p = 0.06$, $p = 0.32$), but it remained significant when each of the other conditions were removed (reverse audio excluded: $F_{(18,324)} = 2.56$, $\eta^2_p = 0.13$, $p < 10^{-3}$; any other condition excluded: $F_{(18,306)} > 1.82$, $\eta^2_p = 0.10$, $p < 0.03$).

Below, we report several additional analyses exploring the fROI-by-condition interaction (or lack thereof). Taken together, these analyses find no evidence that inferior frontal and temporal fROIs have functionally distinct temporal receptive windows.

3.1.2. Controlling for baseline differences in input tracking across fROIs—

Language fROIs differed from one another in their “baseline” input tracking: a one-way ANOVA performed on the intact-story ISCs with fROI (7 levels) as a within-participant

factor revealed a significant main effect ($F_{(6,108)} = 5.13$, $\eta^2_p = 0.22$, $p = 10^{-4}$). To control for these differences, we used the ISCs for the intact story as a “ceiling” against which to “normalize” the ISCs for the other four stimuli. Then, we re-ran the ANOVA testing for fROI-by-condition interaction on the normalized values. Following the patterns observed above, we limited this analysis only to the sentence list and word list conditions, which in our main analyses captured well the characteristics of the full dataset. (We observed that, due to noise in the data, many ISCs for scrambled stimuli exceeded the corresponding ISC for the intact condition and, thus, resulted in a normalized ISC value of 1. These values biased the distribution of ISCs and were difficult to interpret, but the dataset limited to only the sentence list and word lists conditions appeared to lend itself more readily to analysis). As in our main analysis, this test revealed a fROI-by-condition interaction ($F_{(6,108)} = 3.39$, $\eta^2_p = 0.16$, $p = 0.004$) that was accounted for by the MFG fROI (MFG fROI excluded: $F_{(5,90)} = 1.22$, $\eta^2_p = 0.06$, $p = 0.30$; any other region excluded: $F_{(5,90)} > 3.32$, $\eta^2_p > 0.15$, $p < 0.008$). Therefore, it is unlikely that evidence for distinct TRWs across language fROIs was “masked” by regional differences in baseline input tracking.

3.1.3. Testing smaller fROIs—The same results as in the main analysis were obtained when we tested smaller fROIs defined as the top 4% (rather than 10%) of voxels showing the strongest localizer contrast effects within each mask (Fig. 2B): when all 7 fROIs and 5 conditions were included, there was a fROI-by-condition interaction ($F_{(24,408)} = 1.62$, $\eta^2_p = 0.09$, $p = 0.03$), which was no longer significant once the MFG fROI was excluded ($F_{(20,340)} = 1.20$, $\eta^2_p = 0.07$, $p = 0.25$). Similarly, when we tested only the sentence list and word list conditions, there was a significant fROI-by-condition interaction across the 7 fROIs ($F_{(6,108)} = 2.80$, $\eta^2_p = 0.13$, $p < 0.02$) but not across the six fROIs excluding the MFG fROI ($F_{(20,340)} = 1.48$, $\eta^2_p = 0.08$, $p = 0.21$). Thus, the lack of evidence for a fROI-by-condition interaction in our main analysis is unlikely to result from using regions that were too large and grouped together several, functionally distinct sub-regions.

3.1.4. Testing less language-like fROIs—The lack of evidence for a fROI-by-stimulus interaction among core language regions might have resulted from lack of power to detect such interactions. To examine this possibility, we conducted the same analysis on each of several sets of alternative fROIs that, instead of showing the strongest *sentences* > *nonwords* effects, consisted of voxels that showed weaker localizer contrast effects. Specifically, the localizer contrast effects in these voxels were either in the 92–96 percentiles of their respective masks (“second-best” 4%), 88–92 percentiles, 84–92 percentiles, or 80–84 percentiles (“third-”, “fourth-”, and “fifth-best” 4%, respectively). We reasoned that such voxels, which showed less language-like responses, could either be more peripheral members of the language network, or belong to other functional networks that lie in close proximity to language regions (Chein et al., 2002; Fedorenko et al., 2012a; Deen et al., 2015). As such, these regions might differ from one another, and from the core language fROIs, in their integration timescales.

In each of these alternative sets of 7 fROIs, a two-way, repeated-measures ANOVA revealed a fROI-by-condition interaction, indicating that these regions differed from one another in their TRWs (second-best set: $F_{(24,408)} = 1.87$, $\eta^2_p = 0.09$, $p = 0.007$; third-best set: $F_{(24,408)}$

= 1.95, $\eta^2_p = 0.10$, $p = 0.005$; fourth-best set: $F_{(24,408)} = 1.92$, $\eta^2_p = 0.10$, $p = 0.006$; fifth-best set: $F_{(24,408)} = 1.95$, $\eta^2_p = 0.10$, $p = 0.006$). For the fROIs consisting of second-best voxels—voxels that, being within the top 10%, were part of the fROIs used for the main analysis—this interaction was driven by the MFG fROI (fROI-by-condition interaction with the MFG fROI excluded: $F_{(20,340)} = 1.50$, $\eta^2_p = 0.08$, $p = 0.08$). In contrast, for the other sets of fROIs, the interaction remained significant even with the MFG fROI removed, and its effect size descriptively grew as less language-like fROIs were tested (fROI-by-condition interaction with MFG excluded, third-best set: $F_{(20,340)} = 1.78$, $\eta^2_p = 0.09$, $p = 0.02$; fourth-best set: $F_{(20,340)} = 1.96$, $\eta^2_p = 0.10$, $p = 0.008$; fifth-best set: $F_{(20,340)} = 2.35$, $\eta^2_p = 0.12$, $p = 10^{-3}$) (Fig. 2C). These findings demonstrate that our study had sufficient power to detect fROI-by-condition interactions when those exist (e.g., when voxels from nearby functionally distinct networks are examined); such interaction is simply not evident in the core language network, whose regions show indistinguishable TRWs.

3.2. Additional, voxel-based analyses support the main finding

To demonstrate that the lack of evidence for a fROI-by-condition interaction in the core language network was not trivially caused by our choice of the localizer task, we re-computed ISCs using the common, voxel-based approach. First, we observed that the resulting ISCs qualitatively replicate the overall topography of the TRWs reported by Lerner et al. (2011) (Fig. 3A). Namely, large portions of the superior temporal cortex exhibit reliable tracking of all five stimuli, including the reverse audio condition, indicative of a short TRW; middle temporal regions reliably track word lists (as well as less scrambled stimuli) but not the reverse audio, i.e., are sensitive to well-formedness on the timescale of morphemes/syllables or words; extending further in inferior, posterior, and anterior directions, some temporal and parietal regions exhibit longer TRWs and reliably track only stimuli whose structure is well-formed at the level of phrases/clauses or sentences; and in the frontal lobe, voxels exhibit sensitivity to coherence at either the word-, sentence-, or paragraph-level. The broad consistency between this pattern and the previously established pattern of integration timescales indicates that the lack of functional dissociations across core language regions cannot be attributed to fundamental inconsistencies between the current data and those of Lerner et al. (2011).

Next, as a final attempt at uncovering distinct integration timescales across inferior frontal and temporal language regions, we defined an alternative set of fROIs by directly searching for certain TRWs. Recall that, in our descriptive labeling of fROIs in the main analysis (section 3.1.1), two functional profiles were observed: on the one hand, fROIs in the mid-posterior, mid-anterior, and anterior temporal areas exhibited sensitivity to morpheme/syllable or word-level information, with input tracking incurring a significant cost only for the reverse audio. On the other hand, fROIs in the posterior temporal, inferior frontal, and orbital areas exhibited sensitivity to phrase/clause/sentence-level information, with input tracking incurring a cost not only for the reverse audio but also for the word list. Nonetheless, these two profiles did not reliably differ from one another in the main analysis. Now, we tried using a different criterion for defining language regions such that the difference between these two profiles would be maximized. For this purpose, we defined the following fROIs: (i) within each of the former three masks, among those voxels whose ISCs

did not significantly differ between the intact story and word list, we selected the 27 voxels with the biggest difference between ISCs for the intact story and the reverse audio; (ii) within each of the latter three masks, among those voxels whose ISCs did not significantly differ between the intact story and sentence list, we selected the 27 voxels with the biggest difference between ISCs for the intact story and the word list.

These six fROIs were defined based on data from the second half of each stimulus, and we then conducted a two-way, repeated-measures ANOVA to test for a fROI-by-condition interaction in ISCs from the first half of each stimulus (Fig. 3B). The interaction was not significant ($F_{(20,340)} = 1.26$, $\eta^2_p = 0.07$, $p = 0.21$). When testing only the sentence list and word list stimuli, the interaction was very weak ($F_{(5,90)} = 2.35$, $\eta^2_p = 0.12$, $p = 0.047$), especially considering that the choice of which TRW to define in each mask (based on the main analysis) relied on the same functional data tested here (even though, within the current test itself, fROI definition and response estimation were performed on two independent halves of those data). Similar results were obtained when fROIs were defined based on data from the first half of each stimulus and the ANOVA was run on ISCs from the second half (sentence list and word list only: $F_{(5,90)} = 1.54$, $\eta^2_p = 0.08$, $p = 0.19$).

3.3. The core language network as a unified whole occupies a unique stage within a broader cortical hierarchy of integration timescales

The finding that core language regions in inferior frontal and temporal areas show indistinguishable TRWs does not challenge the hypothesis of a broader hierarchy of integration timescales throughout the cortex. Rather, it indicates that core language regions do not occupy multiple, distinct stages within this hierarchy. Yet other functional regions plausibly occupy other stages, some with shorter TRWs than those of the language network (e.g., low-level auditory cortex or speech-perception areas) and others with longer TRWs (e.g., regions engaged in episodic cognition).

To demonstrate this, we first compared the ISCs for the five stimuli averaged across six language fROIs (excluding the MFG fROI) to ISCs from the auditory cortex. A two-way, repeated-measures ANOVA yielded a system (language vs. auditory) by condition interaction ($F_{(4,68)} = 10.53$, $\eta^2_p = 0.38$, $p < 10^{-5}$) (Fig. 4). We followed up on this result with system-by-condition interaction tests that only included the intact story and one other condition. These tests (FDR-corrected for multiple comparisons) revealed that, compared to the core language network, ISCs in the auditory region differed less between the intact story and the word list condition ($F_{(1,18)} = 9.75$, $\eta^2_p = 0.35$, $p = 0.025$), as well as between the intact story and the reverse audio condition ($F_{(1,17)} = 34.11$, $\eta^2_p = 0.67$, $p < 10^{-3}$). In other words, input tracking in the auditory region incurred lower costs for fine-grained scrambling, indicative of a shorter TRW compared to that of the core language network.

Next, we similarly compared ISCs in the language network to those in the episodic network (averaged across left-hemispheric posterior cingulate and temporo-parietal fROIs). Again, we found a network by stimulus interaction ($F_{(4,68)} = 10.82$, $\eta^2_p = 0.69$, $p < 10^{-6}$) (Fig. 4). Follow up interaction tests revealed that, compared to the core language network, ISCs in the episodic network differed more between the intact story and the paragraph list ($F_{(1,18)} = 21.30$, $\eta^2_p = 0.54$, $p = 10^{-3}$), sentence list ($F_{(1,18)} = 15.95$, $\eta^2_p = 0.47$, $p = 0.005$), and

word list ($F_{(1,18)} = 43.68$, $\eta_p^2 = 0.71$, $p < 10^{-4}$) conditions. Input tracking in the episodic network thus incurred higher costs for coarse violations of well-formedness on the timescale of paragraphs and sentences and, further, showed no reliable tracking of word lists. This functional profile is indicative of a longer TRW compared to that of the core language network.

These findings situate the core language network within the context of a cortical hierarchy of integration timescales (Himberger et al., 2018). The common functional profile shared by the inferior frontal and temporal language regions occupies a particular stage within this broader hierarchy, which is located, as expected, downstream from auditory regions and upstream from the episodic network.

4. Discussion

The current study examined how reliably different regions in the core language network track linguistic stimuli that violate well-formedness at various representational grain levels. To this end, we recorded regional time-series of BOLD signal fluctuations elicited by increasingly scrambled versions of a narrated story, and measured the reliability of these fluctuations across individuals to quantify the extent to which they were stimulus-locked (e.g., Hasson et al., 2008). We found that left inferior frontal and temporal language regions all exhibited statistically indistinguishable profiles of sensitivity to linguistic structure at different timescales. Namely, these regions all tracked paragraph lists and sentence lists as reliably as they tracked the intact story, but tracked word lists less reliably, and tracked a reverse audio only weakly or not at all. These findings suggest that language regions integrate information over a common timescale, which is (i) sensitive to structure at the word level or below (e.g., morpheme/syllable), given the increased tracking of the word list compared to the reverse audio; (ii) also sensitive to structure at the phrase/clause or sentence level, given the further increase in tracking of the sentence list compared to the word list; but (iii) not sensitive to information above the sentence level, given no further boosts in tracking of the paragraph list compared to the sentence list. This common profile of information integration provides a novel functional signature of perisylvian, high-level language regions.

We emphasize that our main, null findings of a region-by-condition interaction constitute lack of evidence, and not evidence for a lack of functional dissociations across the core language network. Nevertheless, we extensively tested and rejected alternative explanations for these null results: they are not likely to be accounted for by baseline differences in input tracking across regions, which could have masked differences in TRWs (section 3.1.2); by fROIs being large enough to include—and average across—multiple, functionally distinct sub-regions (section 3.1.3); by lack of power to detect region-by-stimulus interactions in the general cortical areas we focused on (section 3.1.4); or by relying on a task-based, functional localizer to identify participant-specific regions of interest (section 3.2). We therefore conclude that no compelling evidence has been found in favor of a functional dissociation among the regions of the core language network in terms of their integration timescales.

The current results are therefore inconsistent with a division of linguistic labor across the core language network that is topographically organized by integration timescales (cf. Lerner et al., 2011; DeWitt and Rauschecker, 2012; Bornkessel-Schlesewsky et al., 2015; Hasson et al., 2015; Chen et al., 2016; Baldassano et al., 2017; Yeshurun et al., 2017a; Sheng et al., 2018). Instead, they support the hypothesis that inferior frontal and temporal core language regions form a unified whole that occupies a unique stage within a broader cortical hierarchy of temporal integration. In this cortical hierarchy, core language regions follow lower-level auditory (examined here) as well as speech perception regions (Mesgarani et al., 2014; Overath et al., 2015; Poeppel, 2003; Vagharchakian et al., 2012), and precede higher-level associative regions that integrate information over paragraphs and full narratives (Regev et al., 2013; Chen et al., 2016; Margulies et al., 2016; Simony et al., 2016; Yeshurun et al., 2017a; Yeshurun et al., 2017b; Zadbood et al., 2017; Nguyen et al., 2019; Ferstl and von Cramon, 2002; Jacoby and Fedorenko, 2018; Ferstl et al., 2008; Kuperberg et al., 2006; Maguire et al., 1999; Mar, 2011; Yarkoni et al., 2008). The only region within the core language network that might occupy a different stage in this hierarchy is the region that falls within the middle frontal gyrus, which has a somewhat longer integration timescale compared to the rest of the core language network. Given that many of the existing proposals regarding the functional architecture of the language network (for examples, see Friederici, 2002; Hickok and Poeppel, 2004; Ullman, 2004; Grodzinsky and Friederici, 2006; Hickok and Poeppel, 2007; Bornkessel-Schlesewsky and Schlesewsky, 2009; Friederici, 2011, 2012; Poeppel et al., 2012; Price, 2012; Bornkessel-Schlesewsky and Schlesewsky, 2013; Hagoort, 2013) focus on the inferior frontal and temporal regions, to the exclusion of the MFG, we leave further investigation into the relationship between the MFG and the rest of the core language network to future work.

4.1. Evidence for a distributed cognitive architecture of language processing

The finding of a common integration timescale shared across inferior frontal and temporal core language regions constrains comprehension models, challenging the notion of a functional dissociation between processes that either operate on different timescales and/or construct linguistic representations of different grain sizes. Instead, it suggests that linguistic processes at multiple timescales—from syllable/morpheme- or word-level to phrase/clause- or sentence-level—are implemented in neural circuits that are distributed rather than focal and, moreover, overlap with one another and are thus cognitively inseparable.

Nonetheless, our data do not exclude an alternative functional architecture. Recall that the conditions used in our experiment do not map neatly onto traditional psycholinguistic constructs, which are therefore confounded with one another: for instance, the sentence-list condition differs from the word-list condition in the presence of pairwise dependencies between words, multi-word units, phrases, and clauses. Thus, even though the sentence list is tracked more reliably than the word list throughout the core language network, different regions might show this pattern for functionally distinct reasons (e.g., one region might be sensitive to multi-word units, whereas another—to complete phrases). More broadly, inferior frontal and temporal core language regions, which all show the same functional profile in terms of their integration timescale, might each implement a distinct process or set of

processes relevant to such integration (see, e.g., Hagoort and Indefrey, 2014; Brennan et al., 2016) (however, for recent evidence against such a pattern, see Shain et al., 2020).

This alternative view in favor of functional distinctions is inconsistent with much linguistic theorizing (Joshi et al., 1975; Schabes et al., 1988; Goldberg, 1995; Bybee, 1998, 2010; Jackendoff, 2002, 2007; Culicover and Jackendoff, 2005; Wray, 2005; Snider and Arnon, 2012; Langacker, 2008; Christiansen and Arnon, 2017) and empirical behavioral evidence (Clifton et al., 1984; MacDonald et al., 1994; Trueswell et al., 1994; Garnsey et al., 1997; Traxler et al., 2002; Farmer et al., 2006; Reali and Christiansen, 2007; Bradlow and Bent, 2008; Gennari and MacDonald, 2008; Maye et al., 2008; Trude and Brown-Schmidt, 2012; Schmidtke et al., 2014) in favor of a representational continuum that does away with traditionally posited boundaries and extends from phonemes, to morphemes, to words with their syntactic and semantic attributes, to phrase-level constructions and their meanings. A continuous gradient rather than a strict hierarchy of distinct stages is supported by the observation that different language regions exhibited TRWs that, while not statistically distinguishable, nonetheless somewhat descriptively differed from one another.

In addition, our finding of a functional signature distributed across the language network adds to prior neuroimaging studies reporting overlapping and distributed activations across diverse linguistic manipulations (Gernsbacher and Kaschak, 2003; Démonet et al., 2005; Vigneau et al., 2006; Price, 2012). These include manipulations of phonological (Scott and Wise, 2004; Hickok and Poeppel, 2007; Turkeltaub and Coslett, 2010), lexical (Paulesu et al., 1993; Indefrey and Levelt, 2004; Blumstein, 2009; Anderson et al., 2018), syntactic (Caplan, 2007; Bautista and Wilson, 2016; Blank et al., 2016; Pallier et al., 2011), and semantic (Bookheimer, 2002; Thompson-Schill, 2003; Patterson et al., 2007; Binder et al., 2009; Fedorenko et al., 2016, 2020; Mollica et al., 2018; Siegelman et al., 2019) processing. Unlike these traditional, task-based studies, which used controlled manipulations contrived to isolate particular aspects of linguistic processing, the current study employed an alternative approach (Hasson et al., 2004) based on richly structured stimuli in a naturalistic listening paradigm. It therefore importantly complements the prior evidence for a distributed architecture for language processing within which the very same neural circuits support the processing of linguistic units of varying grain size.

4.2. Why functional dissociations among language regions might go undetected

Although we interpret our findings as supporting the distributed implementation of language processing in overlapping neural circuits, they are not inconsistent with some forms of functional dissociations across distinct linguistic mechanisms. Indeed, neuropsychological findings, despite their many inconsistencies, indicate that at least some language regions or pathways may support some linguistic processes and not others, because some individuals with aphasia following brain lesions or degeneration drastically differ from one another in their behavioral symptoms (Caramazza and Coltheart, 2006; Gorno-Tempini et al., 2011). It is thus possible that some fMRI evidence for distributed linguistic processing underestimate a more complex functional architecture within the language network.

Several forms of functional dissociations could have gone undetected in the current study. First, the relatively uncontrolled properties of the naturalistic paradigms and the substantial

differences across stimuli in both the localizer and main tasks might have been unsuitable for detecting subtler linguistic distinctions; and the definition of a single participant-specific fROI in each mask might have compromised our ability to identify distinctions among small regions that lie in close proximity to one another (e.g., Humphries et al., 2005; Hagoort, 2014; Wilson et al., 2018). Nonetheless, whether such distinctions as previously reported in the literature are replicable, and whether they reflect different language-specific functions vs. dissociations between language-specific and other cognitive processes, remains debated (For one such debate, see Dapretto and Bookheimer, 1999; Siegelman et al., 2019; for another, see Frankland and Greene, 2015; Wang et al., 2016; Anderson et al., 2018; see also Vigliocco et al., 2011; Moseley and Pulvermüller, 2014).

Second, due to spatial resolution limits of fMRI, dissociations at the sub-voxel level, including laminar dissociations (e.g., Norris and Polimeni, 2019), could be missed, i.e., functional profiles of distinct neural circuits that are all located within a single voxel would be aggregated together. Decomposing voxel-level BOLD signals into distinct components would require more sophisticated analytic techniques than the ones used here (e.g., Norman-Haignere et al., 2015). Nevertheless, given that the neuropsychological literature has studied many patients with lesions much larger than the spatial grain of fMRI, at least some functional subdivisions within the language network should in principle be detectable across regions rather than within voxels.

Third, the low temporal resolution of the fMRI BOLD signal limits the ability to detect functional distinctions in the time or frequency domains. The neural tracking of linguistic information rapidly evolves over just hundreds of milliseconds (Gross et al., 2013; Ding et al., 2016), but the hemodynamic response to neural activity reaches a peak only after several seconds, smoothing over any putative differences in the timing at which distinct linguistic processes could engage a given region. Similar concerns apply to distinguishing among linguistic operations that operate at distinct frequencies of neural oscillations (e.g., Bastiaansen and Hagoort, 2015). Low temporal resolution would also hinder the detection of differences in the timing at which different regions process the same incoming linguistic stimulus (Dehaene-Lambertz et al., 2006; Stephens et al., 2013; Udden et al., 2019). In particular, given that regions across the core language network are anatomically connected (e.g., Saur et al., 2008) and strongly synchronized in their activity patterns (e.g., Saur et al., 2008; Blank et al., 2014; Braga et al., 2019), information transfer across these regions is likely. Hence, by the time the fMRI BOLD signal is detected, an initially focal neural response might already appear ubiquitous throughout the network. We note, however, that whereas some studies using temporally sensitive methods report multiple, temporally and spatially separable profiles of linguistic integration (Zhang and Ding, 2017), others find that core language regions all exhibit highly similar responses (Fedorenko et al., 2016).

Beyond the methodological limitations of fMRI, studying the division of linguistic labor across the core language network also faces theoretical challenges. As discussed above, traditional distinctions between linguistic constructs (e.g., lexical semantics *vs.* combinatorial syntax) are no longer advocated by many contemporary linguistic and psycholinguistic theories—yet they continue to guide a large portion of neuroimaging studies and neurobiological frameworks (e.g., Ullman, 2004; Friederici, 2012; Bornkessel-

Schlesewsky and Schlesewsky, 2013; Friederici et al., 2017; for discussion, see Fedorenko et al., 2020). Moreover, which functional distinctions should be tested in lieu of the traditionally posited ones remains unclear; in the neuropsychological literature, despite strikingly different behavioral symptoms across some patients, the precise nature of these deficits in cognitive terms and whether they result from damage to different components of the core language network or from damage to functionally distinct networks (the language network vs. nearby, dissociable networks) is still under debate (Caplan et al., 1996, 2007; 2013; Dronkers, 2000; Caramazza et al., 2001; Wilson and Saygin, 2004; Hillis, 2007; Grodzinsky and Santi, 2008). Perhaps, then, in order to better constrain cognitive theories, neuroscientific studies that examine how language processing is divided across distinct circuits should be grounded in contemporary psycholinguistic models and behavioral data.

4.3. A key methodology for neuroimaging studies of language processing

The current study tested whether a previously reported cortical hierarchy of integration timescales (Lerner et al., 2011) functionally corresponded to the core language network, and concluded that language regions all occupy a shared functional stage along this hierarchy. The apparent overlap between the spatial distribution of this hierarchy and the gross topography of the language network is therefore illusory. The key methodological innovation allowing us to demonstrate this point was augmenting the naturalistic paradigm for characterizing temporal receptive windows with a localizer task that identified the core language network in each individual brain. Such participant-specific functional localization established correspondence across brains based on response profiles rather than stereotaxic coordinates, thereby accounting for the substantial inter-individual variability in the precise mapping of function onto macro-anatomy (Duffau, 2017; Frost and Goebel, 2012; Tahmasebi et al., 2012; Vázquez-Rodríguez et al., 2019). Without confounds in the data due to such variability, putative functional dissociations across inferior frontal and temporal language regions in terms of their respective TRWs dissolved, and a common integration timescale was established as a functional signature shared throughout the network.

When inter-subject correlations are instead computed using the common, anatomy-based approach (i.e., on a voxel-by-voxel basis), the resulting functional profiles at the group-level would often not be representative of any individual brain (compare Figs. 3B and 2A). Interpretable ISCs would be obtained only in those stereotaxic coordinates that happened to consistently house the same functional unit across a sufficient number of participants in the sample. In contrast, in functionally heterogeneous cortical areas (Wise et al., 2001; Chein et al., 2002; Fedorenko et al., 2012a; Deen et al., 2015; Braga et al., 2019), a single coordinate could belong to one functional network in some participants but to a second, distinct network in others (Frost and Goebel, 2012; Tahmasebi et al., 2012; Vázquez-Rodríguez et al., 2019; Fedorenko et al., 2020), rendering signal “reliability” (and ISCs as a proxy thereof) an ill-defined concept. For instance, such distinct networks may reliably track respectively independent aspects of a given stimulus, resulting in low ISCs that do not adequately characterize either network. This issue pertains not only to studies of linguistic processing (e.g., Lerner et al., 2014), but to any study based on voxel-wise ISCs including, e.g., the many studies characterizing the episodic network (Honey et al., 2012b; Regev et al., 2013, 2018; Silbert et al., 2014; Simony et al., 2016; Baldassano et al., 2017;

Lahnakoski et al., 2017; Yeshurun et al., 2017a, 2017b), whose location is highly variable across individuals (Braga and Buckner, 2017; DiNicola et al., 2020).

Therefore, we urge researchers relying on ISC measures to confer functional interpretability to their findings by augmenting their approach with a methodology for establishing functional (rather than anatomical) correspondence across brains. For those who take issue with functional localizer tasks, alternative methodologies serving the same purpose are available (e.g., Haxby et al., 2011; Guntupalli et al., 2016; Braga et al., 2019). More generally, as naturalistic stimuli become a core tool in cognitive neuroscience due to their numerous advantages (Maguire, 2012; Ben-Yakov et al., 2012; Sonkusare et al., 2019; Wang et al., 2017; Hasson et al., 2010)—a trend that is celebrated in the current volume—we should not do away with other, well established approaches that have been successful across domains. Rather, a more promising way forward would be to synergistically harness the complementary strengths of multiple paradigms and analytic techniques.

5. Conclusion

As linguistic inputs unfold over time, we integrate them into structured representations that mediate language comprehension. Whereas such integration might proceed hierarchically across the cortex, high-level language regions in the inferior frontal and temporal cortex all occupy a shared functional stage within this hierarchy. We find no evidence for a functional dissociation of integration timescales across these different regions of the core language network. Rather, they all exhibit sensitivity to information that extends from the syllable/morpheme- or word level to the phrase/clause- or sentence level. This finding indicates that the division of linguistic labor across the core language network is not topographically organized according to the grain size of linguistic representations or by distinctions between operations performed on local input *vs.* input spanning larger, more global contexts. Our results are instead more consistent with a spatially distributed set of highly functionally integrated brain regions that implement a language interpretation system where the same mechanisms integrate information in the linguistic input over both relatively short timescales spanning syllables/morphemes and relatively long ones spanning phrases and sentences.

Acknowledgements

We acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research, MIT. We thank Uri Hasson (Psychology Department and the Princeton Neuroscience Institute, Princeton University) and his colleagues for providing the stimuli for the critical task as well as comments on this work. E.F. was supported by NIH awards R00-HD057522, R01-DC016607, and R01-DC016950, a grant from the Simons Foundation to the Simons Center for the Social Brain at MIT, and research funds from the Department of Brain and Cognitive Sciences and the McGovern Institute for Brain Research.

Data/code availability statement

Upon publication of the manuscript, the signal time-courses for all conditions, and from all fROIs, will be published on the Open Science Framework. MATLAB code for reproducing the analyses and bar plots will accompany these data.

References

- Anderson AJ, Lalor EC, Lin F, Binder JR, Fernandino L, Humphries CJ, Conant LL, Raizada RD, Grimm S, Wang X, 2018. Multiple regions of a cortical network commonly encode the meaning of words in multiple grammatical positions of read sentences. *Cerebr. Cortex* 29, 2396–2411.
- Anderson M, Braak CT, 2003. Permutation tests for multi-factorial analysis of variance. *J. Stat. Comput. Simulat* 73, 85–113.
- Andrews-Hanna JR, Reidler JS, Sepulcre J, Poulin R, Buckner RL, 2010. Functional-anatomic fractionation of the brain's default network. *Neuron* 65, 550–562. [PubMed: 20188659]
- Ayyash D, Malik-Moraleda S, Gallee J, Mineroff Z, Jouravlev O, Fedorenko E, (in prep.). The Universal Language Network: A Cross-Linguistic Investigation Spanning 41 Languages and 10 Language Families.
- Baldassano C, Chen J, Zadbood A, Pillow JW, Hasson U, Norman KA, 2017. Discovering event structure in continuous narrative perception and memory. *Neuron* 95, 709–721 e705. [PubMed: 28772125]
- Bastiaansen M, Hagoort P, 2015. Frequency-based segregation of syntactic and semantic unification during online sentence level language comprehension. *J. Cognit. Neurosci* 27, 2095–2107. [PubMed: 26042498]
- Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF, 2003. Voxel-based lesion–symptom mapping. *Nat. Neurosci* 6, 448–450. [PubMed: 12704393]
- Bautista A, Wilson SM, 2016. Neural responses to grammatically and lexically degraded speech. *Lang. Cognit. Neurosci* 31, 567–574. [PubMed: 27525290]
- Ben-Yakov A, Honey CJ, Lerner Y, Hasson U, 2012. Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. *Neuroimage* 63, 501–506. [PubMed: 22813575]
- Benjamini Y, Yekutieli D, 2001. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat* 29, 1165–1188.
- Bhattachali S, Fabre M, Luh W-M, Al Saied H, Constant M, Pallier C, Brennan JR, Spreng RN, Hale J, 2019. Localising memory retrieval and syntactic composition: an fMRI study of naturalistic language comprehension. *Lang. Cognit. Neurosci* 34, 491–510.
- Binder JR, Desai RH, Graves WW, Conant LL, 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebr. Cortex* 19, 2767–2796.
- Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T, 1997. Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci* 17, 353–362. [PubMed: 8987760]
- Blank I, Fedorenko E, 2017. Domain-general brain regions do not track linguistic input as closely as language-selective regions. *J. Neurosci* 37 (41), 9999–10011. [PubMed: 28871034]
- Blank IA, Balewski Z, Mahowald K, Fedorenko E, 2016. Syntactic processing is distributed across the language system. *Neuroimage* 127, 307–323. [PubMed: 26666896]
- Blank IA, Kanwisher N, Fedorenko E, 2014. A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *J. neurophysiol* 112, 1105–1118. [PubMed: 24872535]
- Blumstein SE, 2009. Auditory word recognition: evidence from aphasia and functional neuroimaging. *Lang. Ling. Compass* 3, 824–838.
- Bookheimer S, 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci* 25, 151–188. [PubMed: 12052907]
- Bornkessel I, Zysset S, Friederici AD, von Cramon DY, Schlesewsky M, 2005. Who did what to whom? The neural basis of argument hierarchies during language comprehension. *Neuroimage* 26, 221–233. [PubMed: 15862222]
- Bornkessel-Schlesewsky I, Schlesewsky M, 2009. *Processing Syntax and Morphology: A Neurocognitive Perspective*. Oxford University Press.
- Bornkessel-Schlesewsky I, Schlesewsky M, 2013. Reconciling time, space and function: a new dorsal–ventral stream model of sentence comprehension. *Brain Lang.* 125, 60–76. [PubMed: 23454075]

- Bornkessel-Schlesewsky I, Schlewsky M, Small SL, Rauschecker JP, 2015. Neurobiological roots of language in primate audition: common computational properties. *Trends Cognit. Sci* 19, 142–150. [PubMed: 25600585]
- Bradlow AR, Bent T, 2008. Perceptual adaptation to non-native speech. *Cognition* 106, 707–729. [PubMed: 17532315]
- Braga RM, Buckner RL, 2017. Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. *Neuron* 95, 457–471 e455.
- Braga RM, DiNicola LM, Buckner RL, 2019. Situating the Left-Lateralized Language Network in the Broader Organization of Multiple Specialized Large-Scale Distributed Networks (bioRxiv).
- Branco P, Seixas D, Castro SL, 2020. Mapping language with resting-state functional magnetic resonance imaging: a study on the functional profile of the language network. *Hum. Brain Mapp* 41, 545–560. [PubMed: 31609045]
- Braze D, Mencl WE, Tabor W, Pugh KR, Constable RT, Fulbright RK, Magnuson JS, Van Dyke JA, Shankweiler DP, 2011. Unification of sentence processing via ear and eye: an fMRI study. *Cortex* 47, 416–431. [PubMed: 20117764]
- Brennan J, Nir Y, Hasson U, Malach R, Heeger DJ, Pykkänen L, 2012. Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain Lang.* 120, 163–173. [PubMed: 20472279]
- Brennan JR, Stabler EP, Van Wagenen SE, Luh W-M, Hale JT, 2016. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain Lang.* 157, 81–94. [PubMed: 27208858]
- Buckner RL, Andrews-Hanna JR, Schacter DL, 2008. The brain’s default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci* 1124, 1–38. [PubMed: 18400922]
- Bybee J, 1998. A functionalist approach to grammar and its evolution. *Evol. Commun* 2, 249–278.
- Bybee J, 2010. *Language, Usage and Cognition*. Cambridge University Press, Cambridge, UK.
- Caplan D, 2007. Functional neuroimaging studies of syntactic processing in sentence comprehension: a critical selective review. *Lang. Ling. Compass* 1, 32–47.
- Caplan D, Hildebrandt N, Makris N, 1996. Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain* 119, 933–949. [PubMed: 8673503]
- Caplan D, Michaud J, Hufford R, 2013. Dissociations and associations of performance in syntactic comprehension in aphasia and their implications for the nature of aphasic deficits. *Brain Lang.* 127, 21–33. [PubMed: 24061104]
- Caplan D, Stanczak L, Waters G, 2008. Syntactic and thematic constraint effects on blood oxygenation level dependent signal correlates of comprehension of relative clauses. *J. Cognit. Neurosci* 20, 643–656. [PubMed: 18052788]
- Caplan D, Waters G, DeDe G, Michaud J, Reddy A, 2007. A study of syntactic processing in aphasia I: behavioral (psycholinguistic) aspects. *Brain Lang.* 101, 103–150. [PubMed: 16999989]
- Caramazza A, Capitani E, Rey A, Berndt RS, 2001. Agrammatic Broca’s aphasia is not associated with a single pattern of comprehension performance. *Brain Lang.* 76, 158–184. [PubMed: 11254256]
- Caramazza A, Coltheart M, 2006. Cognitive neuropsychology twenty years on. *Cogn. Neuropsychol* 23, 3–12. [PubMed: 21049319]
- Chein JM, Fissell K, Jacobs S, Fiez JA, 2002. Functional heterogeneity within Broca’s area during verbal working memory. *Physiol. Behav* 77, 635–639. [PubMed: 12527011]
- Chen J, Honey C, Simony E, Arcaro M, Norman K, Hasson U, 2016. Accessing real-life episodic information from minutes versus hours earlier modulates hippocampal and high-order cortical dynamics. *Cerebr. Cortex* 26, 3428–3441.
- Christiansen MH, Aron I, 2017. More than words: the role of multiword sequences in language learning and use. *Top. Cognit. Sci* 9, 542–551. [PubMed: 28503906]
- Christiansen MH, Chater N, 2016. The Now-or-Never bottleneck: a fundamental constraint on language. *Behav. Brain Sci* 39.
- Clifton C, Frazier L, Connine C, 1984. Lexical expectations in sentence comprehension. *J. Verb. Learn. Verb. Behav* 23, 696–708.

- Cordes D, Haughton VM, Arfanakis K, Carew JD, Turski PA, Moritz CH, Quigley MA, Meyerand ME, 2001. Frequencies contributing to functional connectivity in the cerebral cortex in “resting-state” data. *Am. J. Neuroradiol* 22, 1326–1333. [PubMed: 11498421]
- Culicover PW, Jackendoff R, 2005. *Simpler Syntax*. Oxford University Press, Oxford, UK.
- Dapretto M, Bookheimer SY, 1999. Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron* 24, 427–432. [PubMed: 10571235]
- Deen B, Koldewyn K, Kanwisher N, Saxe R, 2015. Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebr. Cortex* 25, 4596–4609.
- Dehaene-Lambertz G, Dehaene S, Anton JL, Campagne A, Ciuciu P, Dehaene GP, D Nghien I, Jobert A, LeBihan D, Sigman M, 2006. Functional segregation of cortical language areas by sentence repetition. *Hum. Brain Mapp* 27, 360–371. [PubMed: 16565949]
- Démonet J-F, Thierry G, Cardebat D, 2005. Renewal of the neurophysiology of language: functional neuroimaging. *Physiol. Rev* 85, 49–95. [PubMed: 15618478]
- Deniz F, Nunez-Elizalde AO, Huth AG, Gallant JL, 2019. The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *J. Neurosci* 39, 7722–7736. [PubMed: 31427396]
- DeWitt I, Rauschecker JP, 2012. Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. Unit. States Am* 109, E505–E514.
- Dick F, Bates E, Wulfeck B, Utman JA, Dronkers N, Gernsbacher MA, 2001. Language deficits, localization, and grammar: evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychol. Rev* 108, 759–788. [PubMed: 11699116]
- Ding N, Melloni L, Zhang H, Tian X, Poeppel D, 2016. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci* 19, 158. [PubMed: 26642090]
- DiNicola LM, Braga RM, Buckner RL, 2020. Parallel distributed networks dissociate episodic and social functions within the individual. *J. neurophysiol* 123, 1144–1179. [PubMed: 32049593]
- Dronkers NF, 2000. The gratuitous relationship between Broca’s aphasia and Broca’s area. *Behav. Brain Sci* 23, 30–31.
- Duffau H, 2017. A two-level model of interindividual anatomo-functional variability of the brain and its implications for neurosurgery. *Cortex* 86, 303–313. [PubMed: 26920729]
- Duffau H, Moritz-Gasser S, Mandonnet E, 2014. A re-examination of neural basis of language processing: proposal of a dynamic hodotopical model from data provided by brain stimulation mapping during picture naming. *Brain Lang.* 131, 1–10. [PubMed: 23866901]
- Farmer TA, Christiansen MH, Monaghan P, 2006. Phonological typicality influences on-line sentence comprehension. *Proc. Natl. Acad. Sci. Unit. States Am* 103, 12203–12208.
- Fedorenko E, (in press). The brain network that supports high-level language processing. . In: Gazzaniga M, Ivery RB, Mangun GR (Eds.), *Cognitive Neuroscience: the Biology of the Mind*. W. W. Norton and Company, New York.
- Fedorenko E, Behr MK, Kanwisher N, 2011. Functional specificity for high-level linguistic processing in the human brain. *Proc. Natl. Acad. Sci. Unit. States Am* 108, 16428–16433.
- Fedorenko E, Blank IA, 2020a. Broca’s area is not a natural kind. *Trends Cognit. Sci* 24, 270–284. [PubMed: 32160565]
- Fedorenko E, Blank IA, Siegelman M, Mineroff Z, 2020. Lack of Selectivity for Syntax Relative to Word Meanings throughout the Language Network, p. 47785 bioRxiv.
- Fedorenko E, Duncan J, Kanwisher N, 2012a. Language-selective and domain-general regions lie side by side within Broca’s area. *Curr. Biol* 22, 2059–2062. [PubMed: 23063434]
- Fedorenko E, Duncan J, Kanwisher N, 2013. Broad domain generality in focal regions of frontal and parietal cortex. *Proc. Natl. Acad. Sci. Unit. States Am* 110, 16616–16621.
- Fedorenko E, Hsieh P-J, 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. [PubMed: 20410363]
- Fedorenko E, Nieto-Castañón A, Kanwisher N, 2012b. Lexical and syntactic representations in the brain: an fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia* 50, 499–513. [PubMed: 21945850]

- Fedorenko E, Scott TL, Brunner P, Coon WG, Pritchett B, Schalk G, Kanwisher N, 2016. Neural correlate of the construction of sentence meaning. *Proc. Natl. Acad. Sci. Unit. States Am* 113, E6256–E6262.
- Fedorenko E, Thompson-Schill SL, 2014. Reworking the language network. *Trends Cognit. Sci* 18, 120–126. [PubMed: 24440115]
- Fedorenko E, Varley R, 2016. Language and thought are not the same thing: evidence from neuroimaging and neurological patients. *Ann. N. Y. Acad. Sci* 1369, 132–153. [PubMed: 27096882]
- Ferstl EC, Neumann J, Bogler C, Von Cramon DY, 2008. The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Hum. Brain Mapp* 29, 581–593. [PubMed: 17557297]
- Ferstl EC, von Cramon DY, 2002. What does the frontomedian cortex contribute to language processing: coherence or theory of mind? *Neuroimage* 17, 1599–1612. [PubMed: 12414298]
- Frank SL, Willems RM, 2017. Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Lang. Cognit. Neurosci* 32, 1192–1203.
- Frankland SM, Greene JD, 2015. An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proc. Natl. Acad. Sci. Unit. States Am* 112, 11732–11737.
- Fridriksson J, den Ouden D-B, Hillis AE, Hickok G, Rorden C, Basilakos A, Yourganov G, Bonilha L, 2018. Anatomy of aphasia revisited. *Brain* 141, 848–862. [PubMed: 29360947]
- Friederici AD, 2002. Towards a neural basis of auditory sentence processing. *Trends Cognit. Sci* 6, 78–84. [PubMed: 15866191]
- Friederici AD, 2011. The brain basis of language processing: from structure to function. *Physiol. Rev* 91, 1357–1392. [PubMed: 22013214]
- Friederici AD, 2012. The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cognit. Sci* 16, 262–268. [PubMed: 22516238]
- Friederici AD, Chomsky N, Berwick RC, Moro A, Bolhuis JJ, 2017. Language, mind and brain. *Nat. Hum. Behav* 1, 713. [PubMed: 31024099]
- Frost MA, Goebel R, 2012. Measuring structural–functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *Neuroimage* 59, 1369–1381. [PubMed: 21875671]
- Garnsey SM, Pearlmutter NJ, Myers E, Lotocky MA, 1997. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *J. Mem. Lang* 37, 58–93.
- Gelman A, 2005. Analysis of variance—why it is more important than ever. *Ann. Stat* 33, 1–53.
- Gennari SP, MacDonald MC, 2008. Semantic indeterminacy in object relative clauses. *J. Mem. Lang* 58, 161–187. [PubMed: 19724662]
- Gernsbacher MA, Kaschak MP, 2003. Neuroimaging studies of language production and comprehension. *Annu. Rev. Psychol* 54, 91–114. [PubMed: 12359916]
- Gloor P, 1997. *The Temporal Lobe and Limbic System*. Oxford University Press, New York, NY.
- Goldberg AE, 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press, Chicago, IL.
- Golland Y, Bentin S, Gelbard H, Benjamini Y, Heller R, Nir Y, Hasson U, Malach R, 2007. Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation. *Cerebr. Cortex* 17, 766–777.
- Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, Ogar JM, Rohrer J, Black S, Boeve BF, Manes F, Dronkers NF, Vandenberghe R, Rascovsky K, Patterson K, Miller BL, Knopman DS, Hodges JR, Mesulam MM, Grossman M, 2011. Classification of primary progressive aphasia and its variants. *Neurology* 76, 1006–1014. [PubMed: 21325651]
- Goucha T, Friederici AD, 2015. The language skeleton after dissecting meaning: a functional segregation within Broca’s Area. *Neuroimage* 114, 294–302. [PubMed: 25871627]
- Grodzinsky Y, Friederici AD, 2006. Neuroimaging of syntax and syntactic processing. *Curr. Opin. Neurobiol* 16, 240–246. [PubMed: 16563739]
- Grodzinsky Y, Santi A, 2008. The battle for Broca’s region. *Trends Cognit. Sci* 12, 474–480. [PubMed: 18930695]

- Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S, 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11, e1001752.
- Guntupalli JS, Hanke M, Halchenko YO, Connolly AC, Ramadge PJ, Haxby JV, 2016. A model of representational spaces in human cortex. *Cerebr. Cortex* 26, 2919–2934.
- Gusnard DA, Raichle ME, 2001. Searching for a baseline: functional imaging and the resting human brain. *Nat. Rev. Neurosci* 2, 685. [PubMed: 11584306]
- Hagoort P, 2013. MUC (memory, unification, control) and beyond. *Front. Psychol* 4, 416. [PubMed: 23874313]
- Hagoort P, 2014. Nodes and networks in the neural architecture for language: broca's region and beyond. *Curr. Opin. Neurobiol* 28, 136–141. [PubMed: 25062474]
- Hagoort P, Indefrey P, 2014. The neurobiology of language beyond single words. *Annu. Rev. Neurosci* 37, 347–362. [PubMed: 24905595]
- Hasson U, Avidan G, Gelbard H, Vallines I, Harel M, Minshew N, Behrmann M, 2009. Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Res.* 2, 220–231. [PubMed: 19708061]
- Hasson U, Chen J, Honey CJ, 2015. Hierarchical process memory: memory as an integral component of information processing. *Trends Cognit. Sci* 19, 304–313. [PubMed: 25980649]
- Hasson U, Malach R, Heeger DJ, 2010. Reliability of cortical activity during natural stimulation. *Trends Cognit. Sci* 14, 40–48. [PubMed: 20004608]
- Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R, 2004. Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. [PubMed: 15016991]
- Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N, 2008. A hierarchy of temporal receptive windows in human cortex. *J. Neurosci* 28, 2539–2550. [PubMed: 18322098]
- Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, Hanke M, Ramadge PJ, 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72, 404–416. [PubMed: 22017997]
- Hickok G, Poeppel D, 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. [PubMed: 15037127]
- Hickok G, Poeppel D, 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci* 8, 393–402. [PubMed: 17431404]
- Hillis AE, 2007. Aphasia progress in the last quarter of a century. *Neurology* 69, 200–213. [PubMed: 17620554]
- Himberger KD, Chien H-Y, Honey CJ, 2018. Principles of temporal processing across the cortical hierarchy. *Neuroscience* 389, 161–174. [PubMed: 29729293]
- Honey CJ, Thesen T, Donner TH, Silbert LJ, Carlson CE, Devinsky O, Doyle WK, Rubin N, Heeger DJ, Hasson U, 2012a. Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* 76, 423–434. [PubMed: 23083743]
- Honey CJ, Thompson CR, Lerner Y, Hasson U, 2012b. Not lost in translation: neural responses shared across languages. *J. Neurosci* 32, 15277–15283.
- Humphreys GF, Hoffman P, Visser M, Binney RJ, Ralph MAL, 2015. Establishing task-and modality-dependent dissociations between the semantic and default mode networks. *Proc. Natl. Acad. Sci. Unit. States Am* 112, 7857–7862.
- Humphries C, Binder JR, Medler DA, Liebenthal E, 2006. Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *J. Cognit. Neurosci* 18, 665–679. [PubMed: 16768368]
- Humphries C, Love T, Swinney D, Hickok G, 2005. Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Hum. Brain Mapp* 26, 128–138. [PubMed: 15895428]
- Indefrey P, Levelt WJ, 2004. The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. [PubMed: 15037128]
- Ivanova AA, Mineroff Z, Zimmerer V, Kanwisher N, Varley R, Fedorenko E, 2019. The Language Network Is Recruited but Not Required for Non-verbal Semantic Processing. *bioRxiv*, p. 696484.

- Jackendoff R, 2002. *Foundation of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, Oxford, UK.
- Jackendoff R, 2007. A parallel architecture perspective on language processing. *Brain Res.* 1146, 2–22. [PubMed: 17045978]
- Jacoby N, Fedorenko E, 2018. Discourse-level comprehension engages medial frontal Theory of Mind brain regions even for expository texts. *Lang. Cognit. Neurosci* 1–17.
- Jones EG, Powell TP, 1970. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain* 93, 793–820. [PubMed: 4992433]
- Joshi AK, Levy LS, Takahashi M, 1975. Tree adjunct grammars. *J. Comput. Syst. Sci* 10, 136–163.
- Jouravlev O, Zheng D, Balewski Z, Pongos ALA, Levan Z, Goldin-Meadow S, Fedorenko E, 2019. Speech-accompanying gestures are not processed by the language-processing mechanisms. *Neuropsychologia* 107132.
- Kandylaki KD, Nagels A, Tune S, Kircher T, Wiese R, Schlesewsky M, Bornkessel-Schlesewsky I, 2016. Predicting “when” in discourse engages the human dorsal auditory stream: an fMRI study using naturalistic stories. *J. Neurosci* 36, 12180–12191.
- Keller TA, Carpenter PA, Just MA, 2001. The neural bases of sentence comprehension: a fMRI examination of syntactic and lexical processing. *Cerebr. Cortex* 11, 223–237.
- Kimura D, Folb S, 1968. Neural processing of backwards-speech sounds. *Science* 161, 395–396. [PubMed: 5667143]
- Koeda M, Takahashi H, Yahata N, Matsuura M, Asai K, Okubo Y, Tanaka H, 2006. Language processing and human voice perception in schizophrenia: a functional magnetic resonance imaging study. *Biol. Psychiatr* 59, 948–957.
- Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI, 2009. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci* 12, 535–540. [PubMed: 19396166]
- Kuperberg GR, Lakshmanan BM, Caplan DN, Holcomb PJ, 2006. Making sense of discourse: an fMRI study of causal inferencing across sentences. *Neuroimage* 33, 343–361. [PubMed: 16876436]
- Lahnakoski JM, ääskeläinen IP, Sams M, Nummenmaa L, 2017. Neural mechanisms for integrating consecutive and interleaved natural events. *Hum. Brain Mapp* 38, 3360–3376. [PubMed: 28379608]
- Langacker RW, 2008. *Cognitive Grammar: A Basic Introduction*. Oxford University Press, Oxford, UK.
- Lerner Y, Honey CJ, Katkov M, Hasson U, 2014. Temporal scaling of neural responses to compressed and dilated natural speech. *J. neurophysiol* 111, 2433–2444. [PubMed: 24647432]
- Lerner Y, Honey CJ, Silbert LJ, Hasson U, 2011. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci* 31, 2906–2915. [PubMed: 21414912]
- MacDonald MC, Pearlmutter NJ, Seidenberg MS, 1994. The lexical nature of syntactic ambiguity resolution. *Psychol. Rev* 101, 676–703. [PubMed: 7984711]
- Maguire EA, 2012. Studying the freely-behaving brain with fMRI. *Neuroimage* 62, 1170–1176. [PubMed: 22245643]
- Maguire EA, Frith CD, Morris R, 1999. The functional neuroanatomy of comprehension and memory: the importance of prior knowledge. *Brain* 122, 1839–1850. [PubMed: 10506087]
- Mahowald K, Fedorenko E, 2016. Reliable individual-level neural markers of high-level language processing: a necessary precursor for relating neural variability to behavioral and genetic variability. *Neuroimage* 139, 74–93. [PubMed: 27261158]
- Mar RA, 2011. The neural bases of social cognition and story comprehension. *Annu. Rev. Psychol* 62, 103–134. [PubMed: 21126178]
- Margulies DS, Ghosh SS, Goulas A, Falkiewicz M, Huntenburg JM, Langs G, Bezgin G, Eickhoff SB, Castellanos FX, Petrides M, 2016. Situating the defaultmode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. Unit. States Am* 113, 12574–12579.
- Matchin W, Hickok G, 2019. *The Cortical Organization of Syntax*.
- Maye J, Aslin RN, Tanenhaus MK, 2008. The weckud wetch of the wast: lexical adaptation to a novel accent. *Cognit. Sci* 32, 543–562. [PubMed: 21635345]

- Meltzer JA, McArdle JJ, Schafer RJ, Braun AR, 2010. Neural aspects of sentence comprehension: syntactic complexity, reversibility, and reanalysis. *Cerebr. Cortex* 20, 1853–1864.
- Menenti L, Gierhan SM, Segaert K, Hagoort P, 2011. Shared language overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychol. Sci* 22, 1173–1182. [PubMed: 21841148]
- Mesgarani N, Cheung C, Johnson K, Chang EF, 2014. Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. [PubMed: 24482117]
- Mesulam M-M, Thompson CK, Weintraub S, Rogalski EJ, 2015. The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia. *Brain* 138, 2423–2437. [PubMed: 26112340]
- Mineroff Z, Blank IA, Mahowald K, Fedorenko E, 2018. A robust dissociation among the language, multiple demand, and default mode networks: evidence from interregion correlations in effect size. *Neuropsychologia* 119, 501–511. [PubMed: 30243926]
- Mirman D, Chen Q, Zhang Y, Wang Z, Faseyitan OK, Coslett HB, Schwartz MF, 2015. Neural organization of spoken language revealed by lesion-symptom mapping. *Nat. Commun* 6.
- Mollica F, Siegelman M, Diachek E, Piantadosi ST, Mineroff Z, Futrell R, Fedorenko E, 2018. High Local Mutual Information Drives the Response in the Human Language Network bioRxiv, 436204.
- Moseley RL, Pulvermüller F, 2014. Nouns, verbs, objects, actions, and abstractions: local fMRI activity indexes semantics, not lexical categories. *Brain Lang.* 132, 28–42. [PubMed: 24727103]
- Nguyen M, Vanderwal T, Hasson U, 2019. Shared understanding of narratives is correlated with shared neural responses. *Neuroimage* 184, 161–170. [PubMed: 30217543]
- Nieto-Castañón A, Fedorenko E, 2012. Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage* 63, 1646–1669. [PubMed: 22784644]
- Nieuwenhuis S, Forstmann BU, Wagenmakers E-J, 2011. Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat. Neurosci* 14, 1105–1107. [PubMed: 21878926]
- Norman-Haignere S, Kanwisher N, McDermott JH, 2015. Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88, 1281–1296. [PubMed: 26687225]
- Norris DG, Polimeni JR, 2019. Laminar (f) MRI: a short history and future prospects. *Neuroimage* 197, 643. [PubMed: 31059800]
- Overath T, McDermott JH, Zarate JM, Poeppel D, 2015. The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nat. Neurosci* 18, 903–911. [PubMed: 25984889]
- Pallier C, Devauchelle A-D, Dehaene S, 2011. Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. Unit. States Am* 108, 2522–2527.
- Patterson K, Nestor PJ, Rogers TT, 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci* 8, 976–987. [PubMed: 18026167]
- Paulesu E, Frith CD, Frackowiak RS, 1993. The neural correlates of the verbal component of working memory. *Nature* 362, 342–345. [PubMed: 8455719]
- Paunov AM, Blank IA, Fedorenko E, 2019. Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *J. neurophysiol* 121, 1244–1265. [PubMed: 30601693]
- Poeppel D, 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun.* 41, 245–255.
- Poeppel D, Emmorey K, Hickok G, Pylkkänen L, 2012. Towards a new neurobiology of language. *J. Neurosci* 32, 14125–14131. [PubMed: 23055482]
- Poldrack RA, 2006. Can cognitive processes be inferred from neuroimaging data? *Trends Cognit. Sci* 10, 59–63. [PubMed: 16406760]
- Poldrack RA, 2011. Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron* 72, 692–697. [PubMed: 22153367]

- Price CJ, 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62, 816–847. [PubMed: 22584224]
- Pritchett BL, Hoeflin C, Koldewyn K, Dechter E, Fedorenko E, 2018. High-level language processing regions are not engaged in action observation or imitation. *J. neurophysiol* 120, 2555–2570. [PubMed: 30156457]
- Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL, 2001. A default mode of brain function. *Proc. Natl. Acad. Sci. Unit. States Am* 98, 676–682.
- Realì F, Christiansen MH, 2007. Processing of relative clauses is made easier by frequency of occurrence. *J. Mem. Lang* 57, 1–23.
- Regev M, Honey CJ, Simony E, Hasson U, 2013. Selective and invariant neural responses to spoken and written narratives. *J. Neurosci* 33, 15978–15988.
- Regev M, Simony E, Lee K, Tan KM, Chen J, Hasson U, 2018. Propagation of information along the cortical hierarchy as a function of attention while reading and listening to stories. *Cerebr. Cortex* bhy282.
- Rolls ET, Joliot M, Tzourio-Mazoyer N, 2015. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *Neuroimage* 122, 1–5. [PubMed: 26241684]
- Saur D, Kreher BW, Schnell S, Kümmerer D, Kellmeyer P, Vry M-S, Umarova R, Musso M, Glauche V, Abel S, 2008. Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. Unit. States Am* 105, 18035–18040.
- Saxe R, Brett M, Kanwisher N, 2006. Divide and conquer: a defense of functional localizers. *Neuroimage* 30, 1088–1096. [PubMed: 16635578]
- Schabes Y, Abeille A, Joshi AK, 1988. Parsing strategies with “lexicalized” grammars: application to tree adjoining grammars. In: *Proceedings of the 12th Conference on Computational Linguistics*, pp. 578–583. Stroudsburg, PA.
- Schmidke DS, Conrad M, Jacobs AM, 2014. Phonological iconicity. *Front. Psychol* 5, 80. [PubMed: 24575062]
- Scott SK, Wise RJ, 2004. The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 92, 13–45. [PubMed: 15037125]
- Scott TL, 2020. *The Neural Basis of Phonological Working Memory*. Boston University.
- Scott TL, Galle J, Fedorenko E, 2016. A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cognit. Neurosci* 8 (3), 167–176. [PubMed: 27386919]
- Shain C, Blank IA, van Schijndel M, Schuler W, Fedorenko E, 2020. fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia* 138, 107307.
- Sheng J, Zheng L, Lyu B, Cen Z, Qin L, Tan LH, Huang M-X, Ding N, Gao J-H, 2018. The cortical maps of hierarchical linguistic structures during speech perception. *Cerebr. Cortex* 29 (8), 3232–3240.
- Siegelman M, Blank IA, Mineroff Z, Fedorenko E, 2019. An attempt to conceptually replicate the dissociation between syntax and semantics during sentence comprehension. *Neuroscience* 413, 219–229. [PubMed: 31200104]
- Silbert LJ, Honey CJ, Simony E, Poeppel D, Hasson U, 2014. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc. Natl. Acad. Sci. Unit. States Am* 111, E4687–E4696.
- Silver NC, Dunlap WP, 1987. Averaging correlation coefficients: should Fisher’s z transformation be used? *J. Appl. Psychol* 72, 146.
- Simony E, Honey CJ, Chen J, Lositsky O, Yeshurun Y, Wiesel A, Hasson U, 2016. Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun* 7.
- Snider N, Arnon I, 2012. A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon. In: *Frequency effects in language*, pp. 127–163.
- Snijders TM, Vosse T, Kempen G, Van Berkum JJA, Petersson KM, Hagoort P, 2009. Retrieval and unification of syntactic structure in sentence comprehension: an fMRI study using word-category ambiguity. *Cerebr. Cortex* 19, 1493–1503.

- Sonkusare S, Breakspear M, Guo C, 2019. Naturalistic stimuli in neuroscience: critically acclaimed. *Trends Cognit. Sci* 23 (8), 699–714. [PubMed: 31257145]
- Stephens GJ, Honey CJ, Hasson U, 2013. A place for time: the spatiotemporal structure of neural dynamics during natural audition. *J. neurophysiol* 110, 2019–2026. [PubMed: 23926041]
- Stowe LA, Broere CA, Paans AM, Wijers AA, Mulder G, Vaalburg W, Zwarts F, 1998. Localizing components of a complex task: sentence processing and working memory. *Neuroreport* 9, 2995–2999. [PubMed: 9804304]
- Tahmasebi AM, Davis MH, Wild CJ, Rodd JM, Hakyemez H, Abolmaesumi P, Johnsrude IS, 2012. Is the link between anatomical structure and function equally strong at all cognitive levels of processing? *Cerebr. Cortex* 22, 1593–1603.
- Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD, 1992. Testing for nonlinearity in time series: the method of surrogate data. *Phys. Nonlinear Phenom* 58, 77–94.
- Thompson-Schill SL, 2003. Neuroimaging studies of semantic memory: inferring “how” from “where”. *Neuropsychologia* 41, 280–292. [PubMed: 12457754]
- Traxler MJ, Morris RK, Seely RE, 2002. Processing subject and object relative clauses: evidence from eye movements. *J. Mem. Lang* 47, 69–90.
- Trude AM, Brown-Schmidt S, 2012. Talker-specific perceptual adaptation during online speech perception. *Lang. Cognit. Process* 27, 979–1001.
- Trueswell JC, Tanenhaus MK, Garnsey SM, 1994. Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. *J. Mem. Lang* 33, 285–318.
- Turkeltaub PE, Coslett HB, 2010. Localization of sublexical speech perception components. *Brain Lang.* 114, 1–15. [PubMed: 20413149]
- Tyler LK, Marslen-Wilson WD, Randall B, Wright P, Devereux BJ, Zhuang J, Papoutsis M, Stamatakis EA, 2011. Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain* 134, 415–431. [PubMed: 21278407]
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M, 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. [PubMed: 11771995]
- Udden J, Hulten A, Schoffelen J-M, Lam N, Harbusch K, van den Bosch A, Kempen G, Petersson KM, Hagoort P, 2019. Supramodal Sentence Processing in the Human Brain: Fmri Evidence for the Influence of Syntactic Complexity in More than 200 Participants. *bioRxiv*, p. 576769.
- Ullman MT, 2004. Contributions of memory circuits to language: the declarative/ procedural model. *Cognition* 92, 231–270. [PubMed: 15037131]
- Vagharchakian L, Dehaene-Lambertz G, Pallier C, Dehaene S, 2012. A temporal bottleneck in the language comprehension network. *J. Neurosci* 32, 9089–9102. [PubMed: 22745508]
- Vandenbergh R, Nobre AC, Price CJ, 2002. The response of left temporal cortex to sentences. *J. Cognit. Neurosci* 14, 550–560. [PubMed: 12126497]
- Vázquez-Rodríguez B, Suarez LE, Markello RD, Shafiei G, Paquola C, Hagmann P, Van Den Heuvel MP, Bernhardt BC, Spreng RN, Misisic B, 2019. Gradients of structure–function tethering across neocortex. *Proc. Natl. Acad. Sci. Unit. States Am* 116, 21219–21227.
- Vigliocco G, Vinson DP, Druks J, Barber H, Cappa SF, 2011. Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neurosci. Biobehav. Rev* 35, 407–426. [PubMed: 20451552]
- Vigneau M, Beaucois V, Herve P-Y, Duffau H, Crivello F, Houde O, Mazoyer B, Tzourio-Mazoyer N, 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30, 1414–1432. [PubMed: 16413796]
- Vul E, Kanwisher N, 2010. Begging the question: the non-independence error in fMRI data analysis. In: Hanson S, Bunzl M.(Eds.), *Foundational Issues for Human Brain Mapping*. MIT Press, Cambridge, MA, pp. 71–91.
- Wang J, Cherkassky VL, Yang Y, Chang K.-m.K., Vargas R, Diana N, Just MA, 2016. Identifying thematic roles from neural representations measured by functional magnetic resonance imaging. *Cogn. Neuropsychol* 33, 257–264. [PubMed: 27314175]

- Wang J, Ren Y, Hu X, Nguyen VT, Guo L, Han J, Guo CC, 2017. Test–retest reliability of functional connectivity networks during naturalistic fMRI paradigms. *Hum. Brain Mapp* 38, 2226–2241. [PubMed: 28094464]
- Whitfield-Gabrieli S, Nieto-Castanon A, 2012. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.* 2, 125–141. [PubMed: 22642651]
- Willems RM, Van der Haegen L, Fisher SE, Francks C, 2014. On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nat. Rev. Neurosci* 15, 193–201. [PubMed: 24518415]
- Wilson SM, Bautista A, McCarron A, 2018. Convergence of spoken and written language processing in the superior temporal sulcus. *Neuroimage* 171, 62–74. [PubMed: 29277646]
- Wilson SM, Saygin AP, 2004. Grammaticality judgment in aphasia: deficits are not specific to syntactic structures, aphasic syndromes, or lesion sites. *J. Cognit. Neurosci* 16, 238–252. [PubMed: 15068594]
- Wise RJS, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA, 2001. Separate neural subsystems within “Wernicke’s area. *Brain* 124, 83–95. [PubMed: 11133789]
- Wray A, 2005. *Formulaic Language and the Lexicon*. Cambridge University Press.
- Yarkoni T, Speer NK, Zacks JM, 2008. Neural substrates of narrative comprehension and memory. *Neuroimage* 41, 1408–1425. [PubMed: 18499478]
- Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Ilei L, Polimeni JR, 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. neurophysiol* 106, 1125–1165. [PubMed: 21653723]
- Yeshurun Y, Nguyen M, Hasson U, 2017a. Amplification of local changes along the timescale processing hierarchy. *Proc. Natl. Acad. Sci. Unit. States Am* 114, 9475–9480.
- Yeshurun Y, Swanson S, Simony E, Chen J, Lazaridi C, Honey CJ, Hasson U, 2017b. Same story, different story: the neural representation of interpretive frameworks. *Psychol. Sci* 28, 307–319. [PubMed: 28099068]
- Zadbood A, Chen J, Leong YC, Norman KA, Hasson U, 2017. How we transmit memories to other brains: constructing shared neural representations via communication. *Cerebr. Cortex* 27, 4988–5000.
- Zhang L, Pylkkänen L, 2015. The interplay of composition and concept specificity in the left anterior temporal lobe: an MEG study. *Neuroimage* 111, 228–240. [PubMed: 25703829]
- Zhang W, Ding N, 2017. Time-domain analysis of neural tracking of hierarchical linguistic structures. *Neuroimage* 146, 333–340. [PubMed: 27856315]

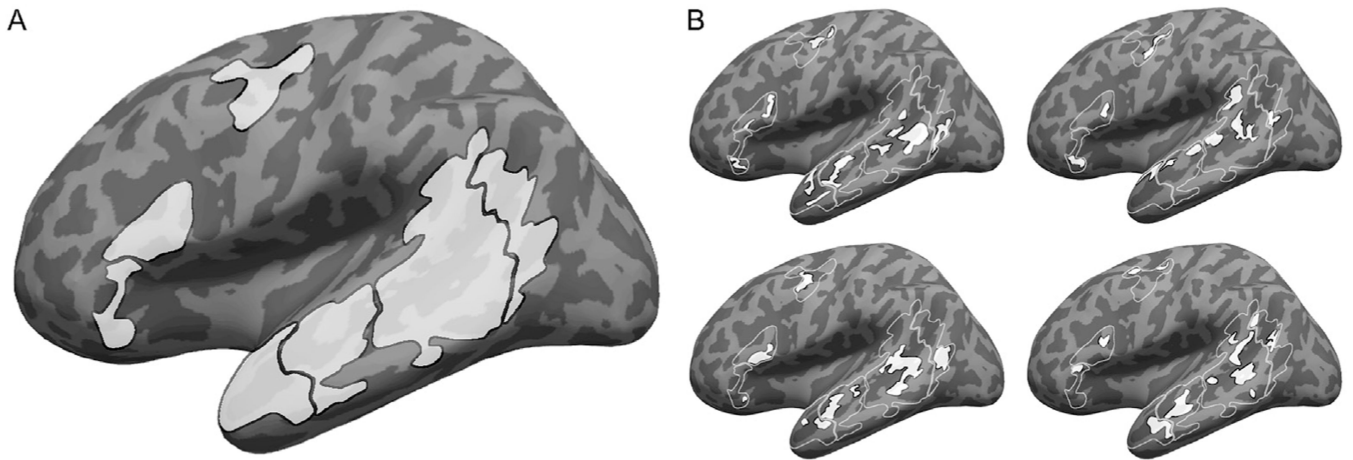


Fig. 1. Defining participant-specific fROIs in the core language network. All images show approximate projections from functional volumes onto the surface of an inflated average brain in common (MNI) space. (A) Group-based masks used to constrain the location of fROIs. These masks were derived from a probabilistic group-level representation of the *sentences* > *nonwords* localizer contrast in a separate sample, following (Fedorenko et al., 2010). Contours of these masks are depicted in white in B. (B) Example fROIs of four participants. Note that, because data were analyzed in volume (not surface) form, some parts of a given fROI that appear discontinuous in the figure (e.g., separated by a sulcus) are contiguous in volumetric space.

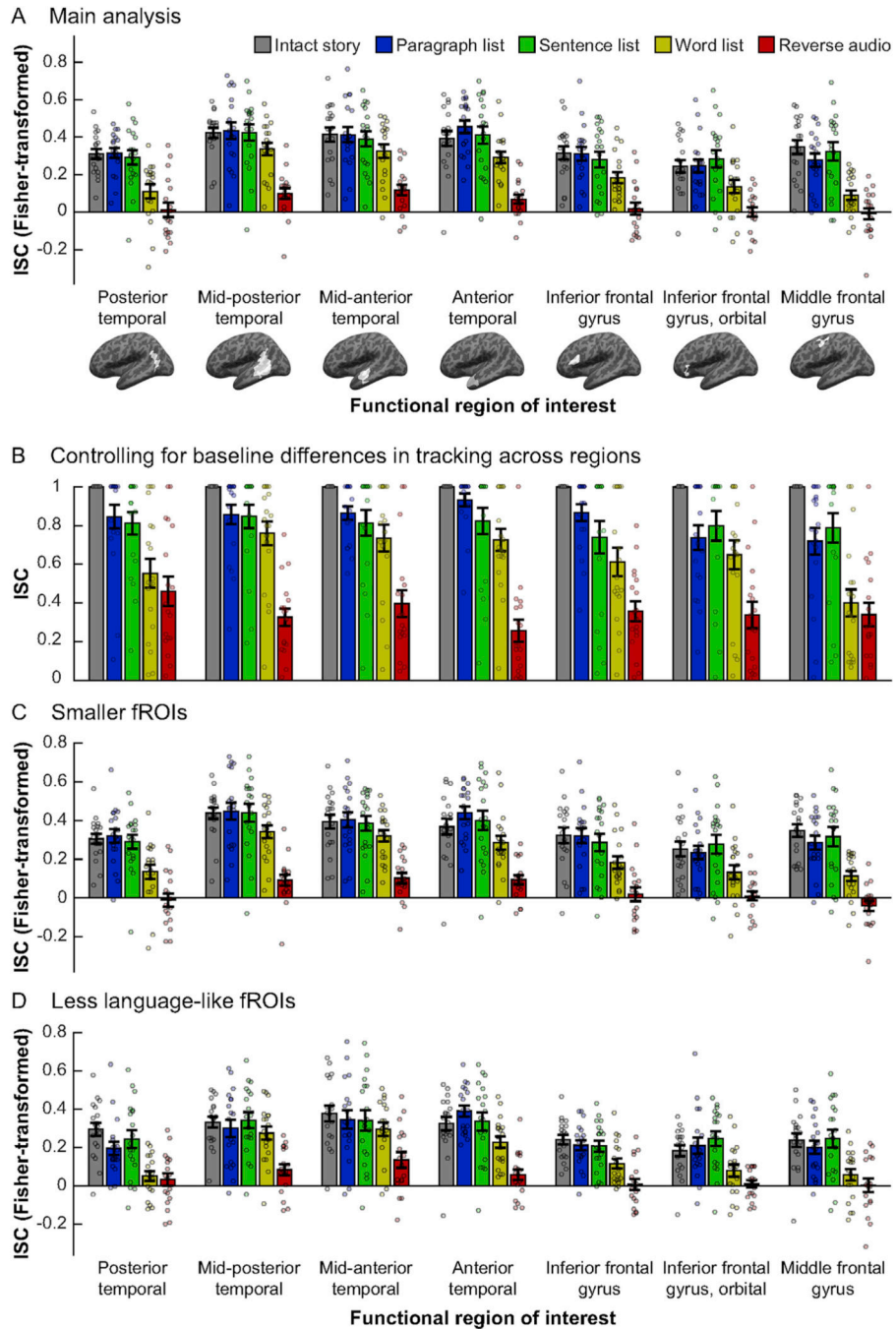


Fig. 2. Comparing temporal receptive windows (TRWs) across core language regions. In each panel, inter-subject correlations (ISCs, a proxy for input tracking reliability; y -axis) are shown for each condition (bar colors) in each fROI (x -axis; brain images depict whole masks, but data are for participant-specific fROIs defined within those masks, as elaborated in Methods). Dots show individual data points, bars show means across the sample, and error-bars show standard errors of the mean. (A) Main analysis. (B) ISCs normalized relative to a “baseline” measure of tracking per fROI, i.e., the intact story condition. (C) ISCs

in smaller fROIs, defined as the top 4% of voxels in each mask that showed the biggest *sentences > nonwords* effect in the localizer task, rather than the top 10% as in (A). (D) fROIs defined as the “fifth-best” 4% of voxels in each mask, i.e., those whose effect size for the localizer contrast was in the 80–84 percentile range. ISCs in all panels, except for panel B, were transformed from the original Pearson’s correlation scale to the Fisher-transformed scale.

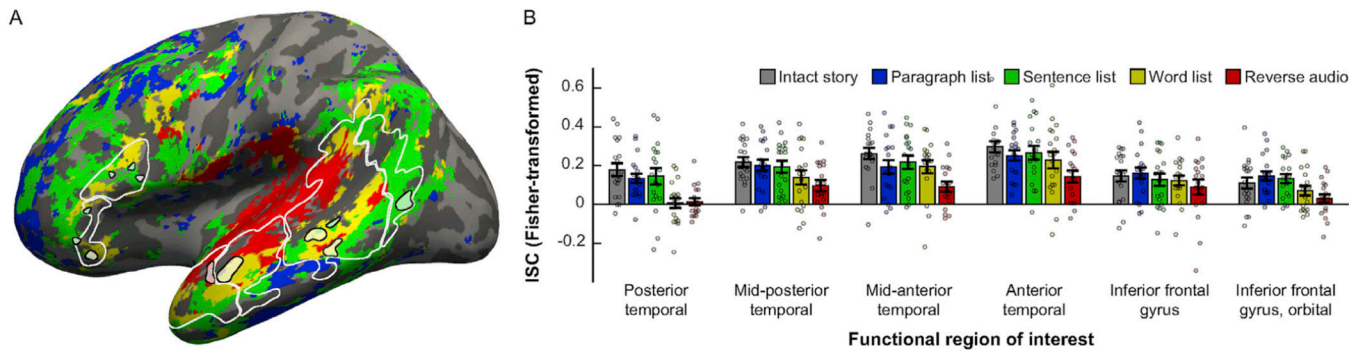


Fig. 3.

Voxel-based analyses. (A) Voxel-based ISCs in the frontal, temporal, and parietal lobes of the left-hemisphere are projected from a functional volume onto the surface of an inflated average brain in common (MNI) space. Each voxel is colored according to its TRW, defined as the most scrambled stimulus that, across participants, was still tracked significantly above chance (following the method in Lerner et al., 2011) (see legend in (B)). Note the progression from a short TRW (significant tracking of all stimuli, labeled in red), around the mid-posterior and mid-anterior temporal lobe, towards longer TRWs (yellow or green) in surrounding temporal areas, and finally to the frontal lobe (yellow, green, or blue). White contours mark the masks that were used for the main analysis (see Fig. 1), excluding the MFG. Small black contours surrounding faint colors mark a set of fROIs that was defined as an alternative to the main fROIs, by choosing the top 27 voxels in each mask showing the strongest pattern consistent with that mask's TRWs as identified in the main analysis (see Fig. 2A). These voxels were chosen based on contrasting ISCs for the intact story to ISCs for each of the other stimuli, and relied on data from only the second half of each stimulus (unlike the coloring scheme in this panel, which is based on data from entire time-series of each stimulus and compared the ISCs of each stimulus against chance). (B) Cross-validated TRWs of the fROIs marked in (A), based on ISCs for data from only the first half of each stimulus. Conventions are the same as in Fig. 2.

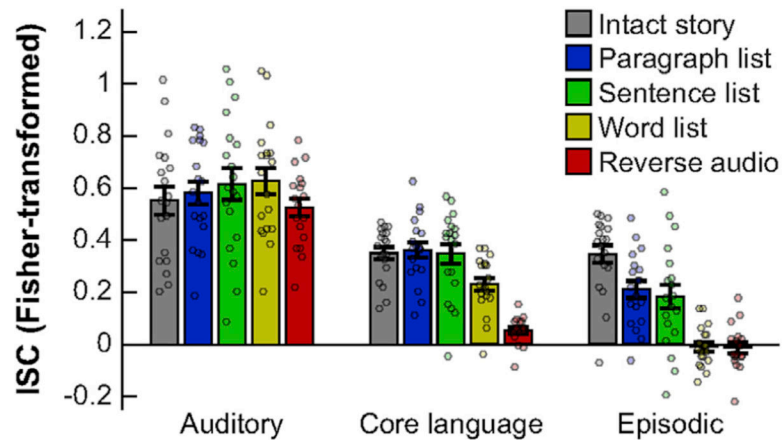


Fig. 4. Demonstrating a hierarchy of TRWs across functionally distinct brain regions. ISCs are shown for a low-level auditory region (around the anterolateral portion of Heschl's gyrus; left), the core language network (averaged across fROIs, excluding the MFG fROI; middle), and a subset of the episodic network (averaged across fROIs in the left posterior cingulate cortex and temporo-parietal junction; right). Conventions are the same as in Fig. 2.