Beyond the "30 Million Word Gap:" Children's Conversational Exposure is Associated with Language-Related Brain Function

Rachel R. Romeo[1,2]*, Julia A. Leonard[2,3], Sydney T. Robinson[2,3], Martin R. West[4], Allyson P. Mackey[2,3,5], Meredith L. Rowe[4], John D. E. Gabrieli[2,3,4]

[1]Division of Medical Sciences, Harvard University

[2]McGovern Institute for Brain Research at the Massachusetts Institute of Technology

[3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

[4]Harvard Graduate School of Education

[5]Department of Psychology, University of Pennsylvania

*Author to whom correspondence should be addressed:

Rachel Romeo, MIT (Office 46-4037), 43 Vassar St, Cambridge, MA 02139;

617-324-6515; rromeo@mit.edu

Abstract

Children's early language exposure impacts their later linguistic skills, cognitive abilities, and academic achievement, and large disparities in language exposure are associated with family socioeconomic status (SES). However, there is little evidence about the neural mechanism(s) underlying the relation between language experience and linguistic/cognitive development. Here, language experience was measured from home audio recordings of 36 SES-diverse 4-6 year-old children. During a story-listening fMRI task, children who had experienced more conversational turns with adults—independent of SES, IQ, and adult/child utterances alone—exhibited greater left inferior frontal (Broca's area) activation, which significantly explained the relation between children's language exposure and verbal skill. This is the first evidence directly relating children's language environments with neural language processing, specifying both environmental and neural mechanisms underlying SES disparities in children's language skills. Furthermore, results suggest that conversational experience impacts neural language processing over and above SES and/or the sheer quantity of words heard.

Keywords: language, socioeconomic status, fMRI, LENA, turn-taking

Introduction

Children's early life experiences during sensitive periods of neural plasticity shape the brain structures and functions underlying their cognitive aptitudes. One critical experience is language exposure. Specifically, the language quantity (e.g., number of words) and quality (e.g., sentence complexity, lexical diversity) that young children hear are the foundation of later language and literacy skills (Hirsh-Pasek et al., 2015; Rodriguez & Tamis-LeMonda, 2011; Rowe, 2012) and non-verbal capacities including executive functioning (Sarsour et al., 2011), math ability (Levine, Suriyakham, Rowe, Huttenlocher, & Gunderson, 2010), and social skills (Connell & Prinz, 2002).

Children's language exposure varies substantially in relation to their socioeconomic status (SES). SES represents the social and economic resources of an individual or group, and children from lower-SES backgrounds on average hear fewer and less complex utterances than their more advantaged peers (Hart & Risley, 1995; Rowe, 2008). In a landmark study, Hart and Risley (1995) estimated that by age 3, children from higher-SES backgrounds had heard 30 million more words than children from lower-SES backgrounds, and other studies report similar trends (Hoff, 2006). Until recently, such studies required time-consuming transcription of parent-child exchanges that limited the amount of data that could be collected. Technological advances now allow for longer, more comprehensive, and less intrusive recordings of naturalistic language exposure. One such device, the Language Environment Analysis (LENA) system, records 16-hour days from the child's perspective and automatically characterizes children's language environments. Studies using LENA have confirmed substantial variation in the amount of language children experience in association with SES (Gilkerson et al., 2017).

This broad or *distal* association between SES and children's language development must be distinguished from the direct or *proximal* association between language exposure and language development (Bronfenbrenner & Morris, 2007). SES is a broad characterization of many correlated factors including income, educational access, other environmental resources, stress, health, and nutrition. Development, however, depends upon specific, proximal factors that directly affect the child, such as immediate language exposure. Indeed, the separability of distal SES from proximal language experience is evident in the considerable variation in early language exposure within each SES band (Gilkerson et al., 2017; Hirsh-Pasek et al., 2015; Rowe, Pan, & Ayoub, 2005; Weisleder & Fernald, 2013). When SES is controlled, children's language exposure remains strongly associated with variation in their language abilities (Rowe, 2012; Weisleder & Fernald, 2013), and differences in exposure partially or fully explain the SES-related gap in language skills (Hoff, 2006).

Despite considerable behavioral research linking children's language exposure to their language abilities, there is currently no evidence about the neural mechanisms underlying this relationship. There is, however, a growing body of evidence that SES disproportionately affects language ability and language neural systems compared to other neurocognitive domains (Farah, 2017). Structurally, lower SES is associated with reduced gray matter in left perisylvian regions underlying phonological, semantic, and syntactic components of language comprehension and production (Noble et al., 2015; Noble, Houston, Kan, & Sowell, 2012), as well as with bilateral occipitotemporal regions involved in reading (Jednorog et al., 2012; Mackey et al., 2015). Additionally, functional neuroimaging with language tasks has revealed SES-related differences in left inferior frontal (Raizada, Richards, Meltzoff, & Kuhl, 2008), superior temporal, and fusiform regions (Noble, Wolmetz, Ochs, Farah, & McCandliss, 2006).

Although these studies provide valuable insight on the relation between brain development and SES, they have not aimed to relate brain measures directly to children's language environments—the proximal factor presumed to directly influence children's linguistic abilities (Noble et al., 2012; Perkins, Finegood, & Swain, 2013). Relating specific and objectively measureable language experiences to brain development is of particular interest because such experiences can become practical and efficacious targets for intervention (Roberts & Kaiser, 2011). Only two neuroimaging studies have related home language experiences to brain functions. One study using fMRI with children ages 8-12 reported a relation between videotaped home language and right prefrontal activation on a complex nonverbal task (Sheridan, Sarsour, Jutte, D'Esposito, & Boyce, 2012). Another study with infants reported a relation between LENA-measured adult word counts and event related potentials (ERPs) to phonetic contrasts in left frontal electrodes (Garcia-Sierra, Ramírez-Esparza, & Kuhl, 2016). However, neither study examined the joint roles of SES and language input in relation to linguistic brain functions.

The present study aims to elucidate how variation in children's natural language experience relates to brain function underlying language processing, and in turn to linguistic abilities. Specifically, we hypothesized that LENA measures of language exposure—over and above SES—would be associated with children's language skills and language-related brain activation, especially in left perisylvian neocortices known to support language.

Method

Participants

Thirty-six children (22 male) aged 4 years, 6 months to 6 years, 10 months (*M* = 5.8 years, *SD* = 0.63 years) and their parents completed this study (see Supplement for justification of sample size). Boys and girls did not significantly differ on any behavioral (all $p > 0.15$), demographic (all $p > 0.33$), language exposure (all $p > 0.76$), or neural measure (maximum $z = 1.2$). Children were native English speakers and typically developing, with no history of premature birth, neurological disorders, developmental delay, speech/language therapy, or grade repetition, and all bilaterally passed a 4 pure-tone hearing screening (0.5 KHz, 1 KHz, 2 KHz, 4KHz) on the day of assessment. Nineteen children were initially assessed and excluded for not meeting these inclusion criteria.

Twenty of the 36 participants additionally participated in a larger randomized controlled intervention study on parenting practices; only their baseline data (before learning of group assignment) was used here. Twenty-seven other children participated but did not have complete data sets, either because they did not complete the home recordings (*n* = 6), did not participate in the fMRI scan (*n* = 11), fell asleep during the fMRI scan (*n* = 3, details below), or exhibited excessive movement during the fMRI scan (*n* = 7, details below). These participants did not differ from the included sample on any behavioral scores, language exposure measures, or SES. All procedures were approved by the Institutional Review Board at the Massachusetts Institute of Technology, and written informed consent was obtained from parents.

Behavioral and Demographic Assessments

Children completed standardized behavioral assessments to characterize verbal and nonverbal cognitive skills (see Supplement for additional info on executive function assessments). These included the Matrix Reasoning, Picture Memory, and Bug Search subtests of the *Wechsler Preschool and Primary Scale of Intelligence* (WPPS-IV) (Wechsler, 2012), the *Peabody Picture Vocabulary Test* (PPVT-4 (Dunn & Dunn, 2007), and the Sentence Comprehension, Word Structure, Formulated Sentences, and Recalling Sentences subtests of the *Clinical Evaluation of Language Fundamentals* (CELF-5) (Wiig, Semel, & Secord, 2013), which together form the CELF-5 Core Language Score (CLS). Age-normed scaled scores from the three WPPSI-IV subtests were averaged to create a "nonverbal composite score." Inclusion criteria required all participants to have nonverbal composite score, PPVT-4 standard score, and CELF-5 CLS greater than or equal to one standard deviation below the mean (16th percentile). Because the CELF-5 only provides age-based norms for children aged 5 years or more, four-year-olds were required to score greater than or equal to the age equivalent for their raw scores on each of the four subtests. Composite Verbal Scores were created by averaging the PPVT-4 standard score and the CELF-5 CLS.

Additionally, parent(s) filled out questionnaires about the child's developmental history and family demographics, including highest level of education obtained by both parents and annual household income. When a father was present in the home, maternal and paternal years of education were averaged to create a parental education metric (1 = high school or less, 2 = some college/associate's degree, 3 = bachelor's degree, 4 = master's/professional degree, 5 = doctoral level degree).

Neuroimaging Data Acquisition

Neuroimaging took place at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research, at the Massachusetts Institute of Technology. First, children were acclimated to the MRI environment and practiced lying still in a mock MRI scanner. Data were then acquired on a 3 Tesla Siemens MAGNETOM Trio Tim scanner equipped for echo planar imaging (EPI; Siemens, Erlangen, Germany) with a 32-channel phased array head coil. An automated scout image was acquired, and shimming procedures were performed to optimize field homogeneity. A whole-head, high-resolution T1-weighted multi-echo MPRAGE structural image was acquired using a protocol optimized for movement-prone pediatric populations (TR = 2530 ms, TE = 1.64 ms/3.5 ms/5.36 ms/7.22 ms, TI = 1400 ms, flip angle = 7°, resolution = 1-mm isotropic). Whole-brain functional images were acquired with a continuous gradient echoplanar T2*-weighted sequence (T2*-weighted images, TR = 2500 ms, TE = 30 ms, flip angle = 90°, bandwidth=2298 Hz/Px, echo spacing= 0.5 ms, 41 transverse slices with FoV =

192 × 192, in-plane resolution of 3 mm × 3 mm). Before each scan, six dummy volumes were acquired and discarded to reach equilibrium, and online prospective acquisition correction (PACE) was applied to the echo planar image sequence throughout the scan.

Functional MRI Task

Children passively listened to short, simple stories derived from the Narrative Language Measures (Petersen & Spencer, 2012), the content of which includes events that young children are likely to be familiar with (e.g., playing games, getting hurt). All stories had consistent narrative structure, word count, and language complexity, and were recorded by a female native-English speaker. A block design paradigm presented fifteen-second long trials consisting of a single story either played normally or played in reverse (backward speech), followed by 5 seconds of silent rest. A third condition (not analyzed here) involved dichotic speech with a different story played in each ear. One run consisted of 6 trials of each condition (18 trials total), such that the run lasted 6 minutes, with condition order pseudo-randomized such that the same condition never repeated twice in a row. Participants were randomly assigned to hear one of two stimulus lists containing all different stories with equal story interest ratings. A female stick figure appeared on a gray screen throughout auditory stimulation to remind children to listen. During the scan, an experimenter stood at the foot of the bore and monitored participants' attentiveness. If the participant closed their eyes for more than 5 seconds, they were considered asleep and their data was discarded (n = 2, mentioned above). Before entering the scanner, children completed a short practice with stories not heard in the scanner, and were required to correctly answer 2 of 4 free-response comprehension questions to ensure familiarity with the task. In the scanner, participants were reminded to listen carefully to the stories to earn prizes upon task completion. Participants were not instructed to memorize the passages, because the goal was to record brain responses during natural language comprehension. Pilot data from children and adults indicated that participants had very low levels of incidental memory for the passages; as such, no post-MRI comprehension/retention test was administered to avoid burdensome additional testing that would be uninformative. All stimuli and scripts are available for download at http://dx.doi.org/10.7910/DVN/DIDBMQ.

Neuroimaging Analysis

Functional MRI data preprocessing and analysis was executed with Nipype, utilizing FSL version 5.0.9 and FreeSurfer version 5.3.0. Functional images were re-aligned to the first volume of the run, co-registered to the corresponding anatomical image (which had been processed and manually edited as necessary in FreeSurfer to ensure correct gray and white matter boundaries), and then to a standard MNI152 template. Functional time-series outliers (global mean intensity > 3 standard deviations, or volume-to-volume motion > 2 mm) were identified by ART and removed from the analysis by adding one regressor per outlier to subject-level general linear models (GLMs). Participants with outliers in more than 20% of volumes were excluded from the study ($n = 7$, mentioned above). Time-series data were high-pass filtered at 120 seconds, spatially smoothed using 6 mm FWHM Gaussian kernel, and convolved with the canonical double-gamma hemodynamic response function (HRF) in FSL, and GLMs were used to create contrast maps for each subject. Subject-level results were combined in mixed effects models using FSL's FEAT with FMRIB's Local Analysis of Mixed Effects (FLAME) stage 1. Results were corrected for multiple comparisons using a conservative cluster-forming threshold of $p < 0.001$, connectivity of 26 (voxels must be connected by at least a point), and a family-wise error rate of $p < 0.05$, and fractionally projected orthogonally to the surface for visualization purposes. Average activations were extracted from subject-level cortical parcellations (according the Desikan-Killiany gyral-based atlas) for mediation analysis.

Home Audio Recordings

Parents were given two LENA Pro digital language processors (DLPs), which are 2-ounce digital recorders that fit in a child's shirt pocket and store up to 16 hours of digitally recorded audio. Parents were instructed to collect full-day recordings from a consecutive Saturday and Sunday, beginning when the child woke up. The average number of days between assessment/MRI and LENA recording was was 8.97 (*SD* = 5.81), with a maximum of 21 intervening days. Upon return of the DLPs, the LENA Pro processing system automatically analyzed the audio and provided estimates of the total number of adult words spoken in the recording (i.e. word tokens), the total number of child utterances, and the total number of adult-child conversational turns, defined as a discrete pair of an adult utterance followed by a child utterance, or vice versa, with no more than 5 seconds pause between the two. Whereas adult words and child utterances are simple linguistic measures, conversational turns incorporate both linguistic information and non-verbal communicative aspects such as temporal contiguity, adult responsiveness, joint social attention, and exchange of communicative information. As such, conversational turns may represent a more holistic measure of interpersonal conversational engagement.

LENA speech-identification algorithms have been determined to be highly reliable, yielding measures approximately 82% accurate for adult speech and 76% accurate for the speech of infants and young children up to 3 years old (Gilkerson et al., 2017; Zimmerman et al., 2009). Although primarily designed to analyze speech of children younger than four years old, the same algorithms were applied to recordings from all participants, such that any potential inaccuracies would be consistent. Running totals for each speech category were calculated for each consecutive 60 minutes across the two days in 5 minute increments (e.g., 7:00 AM – 8:00 AM, 7:05 AM – 8:05 AM, etc.), and the per-participant highest hourly total of adult words, child utterances, and conversational turns were separately extracted for further analysis. This metric helped minimize differences in daily totals due solely to different recording lengths and/or loud activities that may have masked speech and misrepresented language input. It also attempted to reduce the amount of "overheard speech" that was not child-directed, since peak language periods are shown to be more similar to engaged structured play situations (Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017). Such measures of peak naturalistic observations are consistent with other studies utilizing LENA (Garcia-Sierra et al., 2016; Ramírez-Esparza, García-Sierra, & Kuhl, 2014).

Results

Behavioral Results

All data (to the extent that they are available to share) are freely available for download at http://dx.doi.org/10.7910/DVN/DIDBMQ. Children's verbal and nonverbal ability, according to standardized assessments, ranged from low average to above average (verbal composite standard score: Range = 86-139, M = 114, SD = 15; nonverbal composite scaled score: Range = 7.3-14.7, M = 10.6, SD = 2.1). Parental education ranged from partial high school to doctorate level degrees (M = some college), and familial income ranged from $6,000 to $250,000 per year, with median of $85,500 per year, consistent with the median familial income in Massachusetts of $90,590. Parental education, but not income, was positively correlated with children's nonverbal ability (education: $r = 0.34$, 95% CI = [.02, .67], $p < 0.05$; income: $r = 0.11$, p = n.s.; Fig 1a). Although both education and income were correlated with children's verbal ability (education: $r = 0.69$, 95% CI = [.44, .94], $p < 0.00001$; income: $r = 0.48$, 95% CI = [.17, .79], $p < 0.01$; Fig 1b), linear regression revealed that income predicted no unique variance in child verbal ability after accounting for parental education (education: $\beta = 8.25$, 95% CI = [4.07, 12.44], $p < 0.001$, income: β = < .01, p = n.s.).

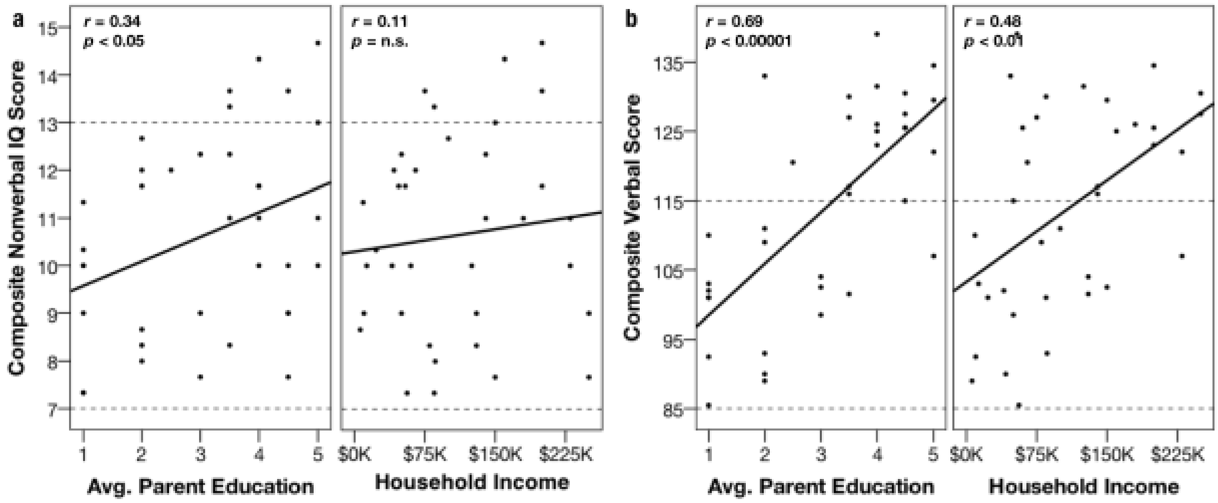

Fig. 1. Scatterplots of composite (a) nonverbal and (b) verbal scores as functions of parent education level (mother and father averaged) and household income. Standardized nonverbal assessments evaluated fluid reasoning, nonverbal working memory, and processing speed. Standardized verbal assessments evaluated vocabulary, receptive and expressive morphosyntax, and verbal working memory skill. Dotted lines indicate the average range of scores (within 1 standard deviation of population mean).

There was great individual variability in language exposure measures, including the number of adult words per peak hour (M = 4260, SD = 1225, range = 1953-6991), the number of child utterances per hour (M = 743, SD = 261, range = 300-1275), and the number of conversational turns per hour (M = 181, SD = 56, range = 86-330). Higher parental education and income correlated significantly with more adult words (education: r = 0.41, 95% CI = [.09, .73], $p < 0.05$; income: r = 0.39, 95% CI = [.06, .71], $p < 0.05$) and more conversational turns (education: r = 0.34, 95% CI = [.02, .67], $p < 0.05$; income: r = 0.37, 95% CI = [.04, .69], $p < 0.05$; Fig. 2), but neither SES measure was significantly correlated with child utterances (education: r = 0.25, p = n.s.; income: r = 0.24, p = n.s.). If these peak-hour measures are extrapolated, children in the top and bottom SES quartiles would experience an annual adult word gap of 5 million words, which could accumulate to approximately 30 million words by age of enrollment in this study, similar to the gap originally reported by Hart and Risley (1995). However, SES only explained a moderate share of the variability in language exposure (11-17%), indicating that there was wide variability of language exposure within families of similar SES.

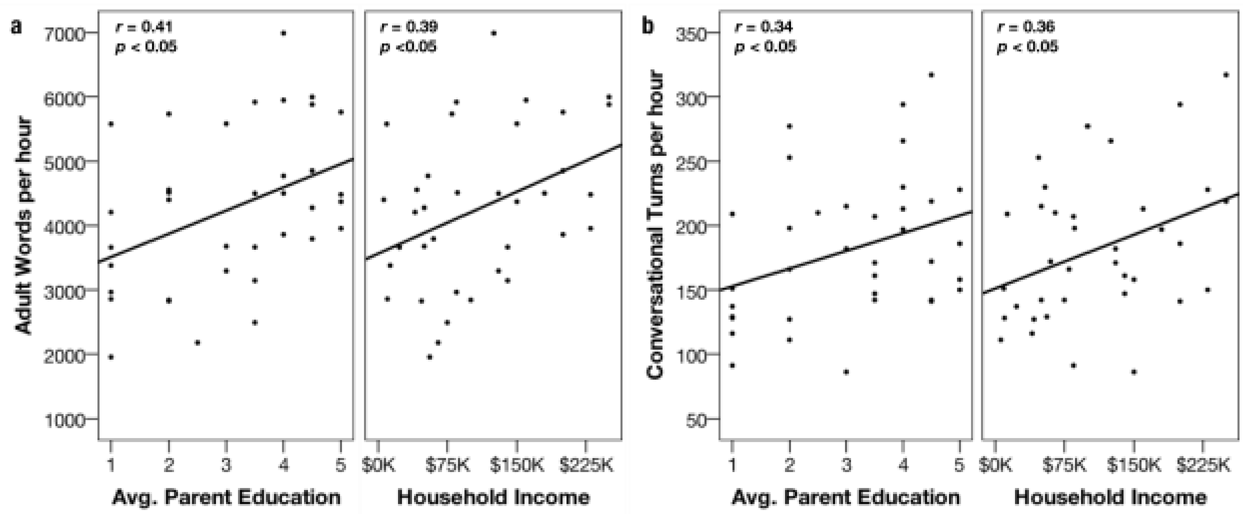

Fig. 2. Scatterplots of peak hourly (a) adult words and (b) conversational turns as functions of parent education level (mother and father averaged) and household income.

All three measures of language experience correlated with children's scores on behavioral language assessments, although conversational turns most strongly predicted the verbal composite score (conversational turns: r = 0.51, 95% CI = [.022, .81], p < 0.001; adult words: r = 0.36, 95% CI = [.04, .69], p < 0.05; child utterances: r = 0.34, 95% CI = [.01, .66], p < 0.05). Multiple regression models were constructed to predict verbal composite scores as a function of parental education, family income, and each of the three language experience measures. In all three models, parental education significantly predicted verbal scores [all β > 7.70, p < 0.001, partial r > 0.55] whereas income did not [all β < 0.1]. Only conversational turns significantly predicted additional variance in verbal scores after education and income were partialled out (β = 0.09, 95% CI = [.02, .16], p = 0.01, partial r = 0.43, R^2 change = 0.10; Fig. 3). Thus, children's composite verbal score increased by one point for every additional 11 conversational turns experienced per hour, independent of SES. The relation between conversational turns and verbal scores remained significant (all $\beta > 0.08$, $p < 0.05$) when adult words and/or child utterances was added to the model, suggesting that conversational turns was not just a proxy for adult speech or child talkativeness. Furthermore, a bootstrap mediation analysis revealed that the number of conversational turns significantly mediated the relationship between parental education and verbal composite scores (indirect effect = 1.16, 95% CI = [0.22, 2.92], indirect/total effect = 0.16), such that variation in conversational turns could account for 16% of the total relationship between parental education and children's verbal scores.

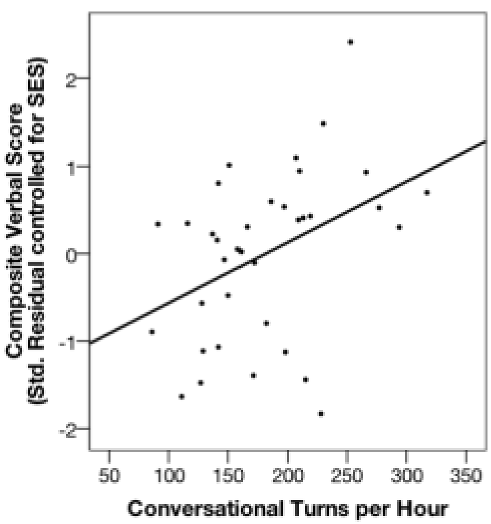

Fig. 3. Relationship between children's composite verbal score (controlled for parent education level and income) and the number of hourly conversational turns.

Neuroimaging Results

The contrast of interest was activation during the comprehensible forward speech condition versus the incomprehensible backward speech condition, which yields activation specific to higher-level language processing involved in comprehending heard stories, roughly controlling for auditory characteristics. As a group, this task yielded significant activation along bilateral superior temporal sulci (STS), with a leftward lateralization (Fig. 4, Table S1); in the left hemisphere, a cluster extended from the temporal pole to supramarginal/angular gyri, while in the right hemisphere, a cluster was restricted to the anterior portion of the STS.

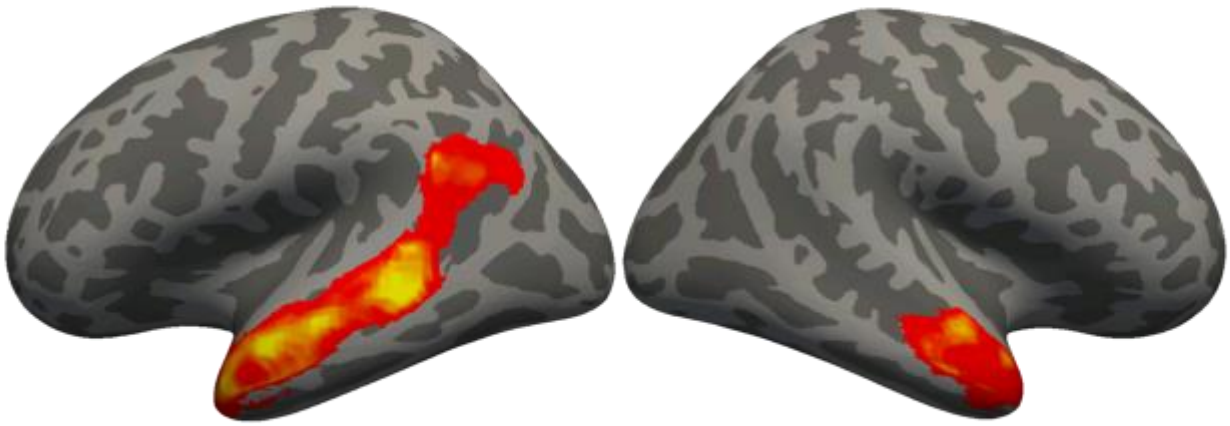

Fig. 4. Regions where activation was significantly greater while listening to forward speech versus backward speech, averaged across all participants. Clusters include the whole of the left superior temporal sulcus and the anterior portion of the right superior temporal sulcus.

Whole brain correlations with the three LENA measures were conducted to detect individual differences in activation related to language exposure. While there were no significant correlations with the number of adult words or child utterances, the number of conversational turns correlated positively with activation in a single cluster (Fig. 5a, Table S2, 766 total voxels) spanning left pars triangularis (Brodmann area 45) extending into pars opercularis (Brodmann area 44), which together comprise "Broca's area." This cluster remained significantly correlated with conversational turns after controlling for parental education and income (Fig. 5b), verbal and nonverbal composite scores (Fig. 5c), adult words and child utterances counts (Fig. 5d), or all of these covariates together (Fig. 5e), indicating that this relationship was not driven simply by any of these factors. In other words, the more conversational turns a child experienced, the greater their activation in Broca's area during language processing, independent of the child's SES, cognitive ability, or sheer numbers of adult words and child utterances. There were no clusters exhibiting significant correlations with any demographic variables (age, gender, parent education, income) or cognitive (verbal, nonverbal scores) variables.

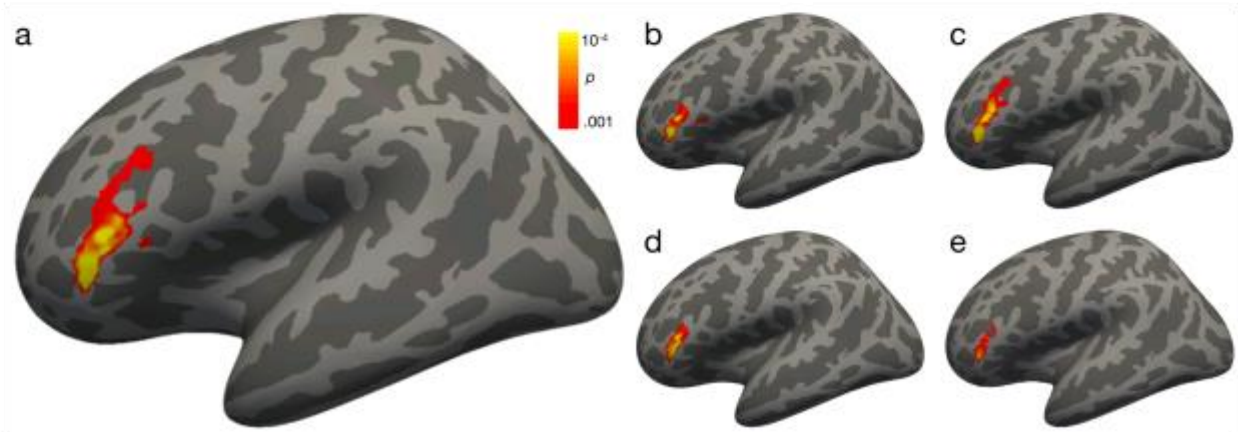

Fig 5. Correlations between activation during language processing and the number of hourly conversational turns children experienced. **(a)** Zero-order correlation between the number of conversational turns and activation in the forward > backward speech contrast. Correlations remained significant when controlling for **(b)** parental education and income, **(c)** verbal and nonverbal assessment scores, **(d)** individual numbers of adult words and child utterances, or **(e)** all of these covariates.

We then asked if Broca's area activation helped explain the relation between children's language exposure and verbal scores. The magnitudes of children's Broca's area activations, (averaged over anatomically-defined opercular and triangular regions, as shown in Fig. 6) significantly mediated the relation between the number of conversational turns and verbal composite scores (indirect effect = 0.065, 95% CI = [0.02, 0.11], indirect/total effect = 0.48), rendering the relation between conversational turns and verbal scores insignificant. This suggests that conversational turns may support children's verbal skills in part by influencing Broca's area activation during language processing. Further, this neural pattern explained 48% of the relation between children's conversational turns and their verbal scores.

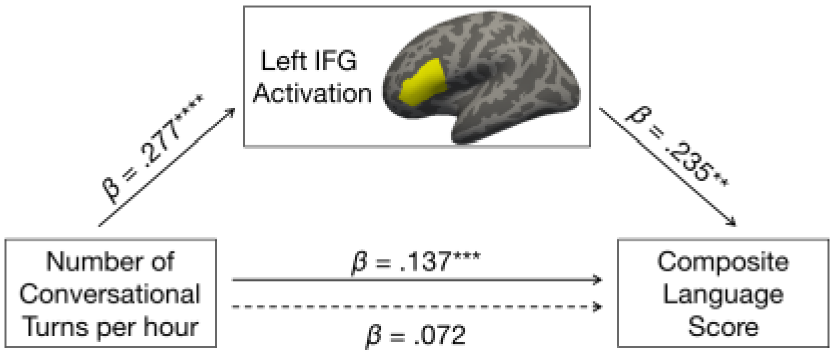

Fig. 6. Mediation model showing the effect of conversational turns on language assessment scores as mediated by activation in the left inferior frontal gyrus, shaded in yellow. Activation significantly mediated the relation between the number of conversational turns children experience and their language scores. Solid arrows represent direct paths, whereas the dotted arrow represents the indirect (mediated) path. β coefficients represent unstandardized regression coefficients. **p* < 0.05, ***p* < 0.01, ****p* < 0.001

Finally, conversational turns and Broca's activation jointly mediated the relationship between parent education and children's language scores, (indirect effect = 1.69, 95% CI = [0.24, 3.75], indirect/total effect = 0.23), indicating that conversational turns and Broca's activation during language processing could account for 23% of the total relationship between SES and children's language skills.

Discussion

This study provides the first evidence of the neural activation patterns underlying the relation between children's early language exposure and verbal skills. Using at-home, real-world audio recorders, we replicated behavioral findings that higher SES is correlated with both greater language experience and verbal abilities in children ages 4-6 years. Specifically, it was the number of conversational turns between children and adults (and not the sheer number of adult words) that significantly mediated the SES-verbal ability relationship. Further, neuroimaging revealed a neural mechanism by which language experience may influence brain development; namely, children who experienced more conversational turns exhibited greater activation in left inferior frontal regions ("Broca's area") during language processing, which explained nearly half the relationship between children's language exposure and verbal abilities. Finally, conversational turns and Broca's activation jointly mediated the relationship between SES and children's language abilities, demonstrating both environmental and neural mechanisms underlying SES disparities in early language skills.

These findings are consistent with evidence that *qualitative* aspects of children's language experience (such as turn-taking) may have a greater impact on language development than sheer *quantitative* measures (Hirsh-Pasek et al., 2015; Zimmerman et al., 2009). While the conversational turn count likely includes more child-directed speech than the adult word count (which also includes any "overheard" speech), it is unlikely that the quantity of child-directed speech alone explains the significance of the conversational turn measure. Studies of child-directed speech suggest that contiguity (temporal connectedness) and contingency (contextual relevancy) with children's utterances are critical for word-learning (Roseberry, Hirsh-Pasek, & Golinkoff, 2014), and that the fluency, connectedness, and joint engagement of communication predict later language skills over and above the number of adult words (Hirsh-Pasek et al., 2015). In fact, conversational turns fully explain the effect of adult words on 2-to-48-month-old children's language skills (Zimmerman et al., 2009). The present results extend the importance of conversational turns to language skills at age 6, suggesting a continued role for this essentially social aspect of language development.

Conversational turns may be particularly important for language development because they provide increased opportunities for children to practice language and receive feedback from adults. Furthermore, this creates a feedback loop to help adults hone their own speech to the optimal complexity to best support children's language development (Zimmerman et al., 2009). While it is possible that children with better language abilities may better engage in these conversations, child utterances had the weakest relation to language scores and brain functions, suggesting that the strong effect of conversational turns is not simply a reflection of more talkative children. More broadly, the importance of conversational turns supports theories that language development crucially relies on social interaction and social neural circuitry (Kuhl, 2007) and that pre-linguistic communicative turn-taking was essential for the evolution of language (Levinson, 2016).

The present study is the first to provide evidence of a localized (left inferior frontal) neural mechanism that underlies the relation between children's direct language exposure and language processing. This is consistent with findings that language input is related to infants' ERP responses in left frontal regions during a phonological task (Garcia-Sierra et al., 2016). Thus, linguistic experience appears to have a particular influence on language processes in left prefrontal cortex, beginning in infancy and continuing through early childhood.

The finding that participants as a group yielded left-lateralized superior temporal activation is likely indicative of a relative invariance in activation related to the acoustic/sub-discourse aspects of language. However, *variation* in participants' language experience correlated exclusively with activation in Broca's area. The localization of this brain-behavior relationship may be related to the nature of conversational turns as a higher-level, supralexical language process. Although Broca's area is classically associated with speech production, research suggests it plays a much broader role in both receptive and expressive language processing, as well as a variety of non-linguistic functions. The specific role of Broca's area in passive language comprehension is still a matter of debate, although it may function as a convergence zone, in which small, independent elements of language (e.g., phonemes, words) are unified into a coherent overall representation (Hagoort, 2014). The present functional task—listening to meaningful, connected stories—requires integration across phonological, semantic, and syntactic units; thus, greater activation in Broca's area may represent a deeper engagement with the linguistic structure of the stories. Alternatively, regions of Broca's area also support several domain-general functions, including action perception, working memory, and executive functioning/cognitive control (Fedorenko, Duncan, & Kanwisher, 2012); by this view, greater activation could represent a neural representation of the speaker's/characters' movements, and/or relating current verbal information to recently heard sentences/stories. Conversational experience could plausibly contribute to either or both neural systems, and future studies are needed to delineate the precise cognitive process(es) associated with language exposure.

Several limitations of the present study are noted. To study typical development, children with language disorders/delays or language scores below the 16th percentile were excluded. Given the strong relation between SES and language scores, this may have disproportionately excluded lower-SES children, which some argue may itself be considered a "learning disability" (Ryan, 2013). As such, future studies should delineate the generalization of these findings to children with a greater variety of language abilities. Additionally, participants' young age required minimization of in-scanner tasks; as such, the functional task was passive in nature. Although children were required to demonstrate listening comprehension before entering the scanner, monitored for alertness during scanning, and incentivized to listen closely, children could have varied in their level of task engagement. However, this is unlikely to wholly account for activation differences, because there were no temporal-lobe differences in relation to language experience and because this task has revealed robust perisylvian activation even in young, sleeping children (Redcay, Haist, & Courchesne, 2008). Nevertheless, any functional activation is constrained by the nature of the task and material used in an experiment, and further studies will be needed to characterize the scope and limits of the present findings.

Finally, while LENA provides immense, naturalistic data on the quantity of speech experienced, it does not parse *what* is said, and thus provides little information about other qualitative aspects of language, such as lexical diversity and grammatical complexity. Future studies should determine the precise relation between conversational quantity and quality on brain and language development.

Although it has been theorized that the home language environment underlies the link between SES and the structure and function of canonical language-related brain regions (Noble et al., 2012; Perkins et al., 2013), this is the first study to reveal a direct relation between a specific aspect of language exposure, namely conversational turns, and brain function during language processing. While causation cannot be implied, results suggest that early language exposure, a proximal aspect of children's environment, may alter the way in which their brains process language. These findings also have clear practical implications. While many early intervention programs aim to increase the amount of language parents address *to* their children, these findings suggest programs should also encourage parents to talk *with* their children by engaging in more interactive, back-and-forth conversation (Leech, Wei, Harring, & Rowe, in press; McGillion, Pine, Herbert, & Matthews, 2017). Future longitudinal studies may determine if increasing the number of conversational turns affects the neural patterns supporting language processing, and if there is a critical/sensitive period for such neural changes. Nevertheless, the present study provides initial information on the neural mechanisms underlying the link between children's linguistic exposure and their language development.

Author Contributions

R. R. Romeo and J. D. E. Gabrieli developed the study concept. R. R. Romeo, J. A. Leonard, A. P. Mackey, M. L. Rowe, and J. D. E. Gabrieli designed the study. R. R. Romeo, J. A. Leonard, and S. T. Robinson collected the data. R. R. Romeo performed the data analysis and interpretation under the supervision of J. D. E. Gabrieli and M. L. Rowe. R. R. Romeo drafted the manuscript, and J. A. Leonard, M. R. West, A. P. Mackey, M. L. Rowe, and J. D. E. Gabrieli provided critical revisions. All authors approved the final version of the manuscript for submission.

Acknowledgements

We thank the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research (MIT); Atshusi Takahashi, Steve Shannon, and Sheeba Arnold for data collection support; Kelly Halverson, Emilia Motroni, Lauren Pesta, Veronica Wheaton, and Christina Yu for assistance in administering behavioral assessments; Megumi Takada for help with data collection/organization; Anne Fernald for insight on LENA data analysis; Joshua Segaran and Hannah Grotzinger for MRI quality assurance; Tyler Perrachione for thoughtful conversations; Andrea Imhof for manuscript comments; and Transforming Education, John Connolly and Glennys Sanchez from 1647 Families plus Ethan Scherer from the Boston Charter Research Collaborative for extensive recruitment support.

Funding

Research was funded by the Walton Family Foundation (to M.R.W.), National Institute of Child Health and Human Development (F31HD086957 to R.R.R.), Harvard Mind Brain Behavior Grant (to R.R.R.), and a gift from David Pun Chan.

References

Bronfenbrenner, U., & Morris, P. A. (2007). The bioecological model of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology* (6 ed., Vol. 1: Theoretical models of human development, pp. 793-828). Hoboken, NJ: Wiley.

Connell, C. M., & Prinz, R. J. (2002). The impact of childcare and parent–child interactions on school readiness and social skills development for low-income African American children. *J Sch Psychol, 40*(2), 177-193.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test* (4th ed.). Bloomington, MN: Pearson.

Farah, M. J. (2017). The Neuroscience of Socioeconomic Status: Correlates, Causes, and Consequences. *Neuron, 96*(1), 56-71.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Curr Biol, 22*(21), 2059-2062.

Garcia-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2016). Relationships between quantity of language input and brain responses in bilingual and monolingual infants. *Int J Psychophysiol, 110*, 1-17.

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Kimbrough Oller, D., et al. (2017). Mapping the early language environment using all-day recordings and automated analysis. *Am J Speech Lang Pathol, 26*(2), 248-265.

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Curr Opin Neurobiol, 28*, 136-141.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: P.H. Brookes.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff , R. M., Pace, A., et al. (2015). The contribution of early communication quality to low-income children's language success. *Psychol Sci, 26*(7), 1071-1083.

Hoff, E. (2006). How social contexts support and shape language development. *Dev Rev, 26*(1), 55-88.

Jednorog, K., Altarelli, I., Monzalvo, K., Fluss, J., Dubois, J., Billard, C., et al. (2012). The influence of socioeconomic status on children's brain structure. *PLoS One, 7*(8), e42486.

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Dev Sci, 10*(1), 110-120.

Leech, K. A., Wei, R., Harring, J., & Rowe, M. L. (in press). A brief parent-focused intervention to improve preschoolers' conversational skills and school readiness. *Dev Psychol*.

Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., & Gunderson, E. A. (2010). What counts in the development of young children's number knowledge? *Dev Psychol, 46*(5), 1309-1319.

Levinson, S. C. (2016). Turn-taking in human communication--Origins and implications for language processing. *Trends Cogn Sci, 20*(1), 6-14.

Mackey, A. P., Finn, A. S., Leonard, J. A., Jacoby-Senghor, D. S., West, M. R., Gabrieli, C. F., et al. (2015). Neuroanatomical correlates of the income-achievement gap. *Psychol Sci, 26*(6), 925-933.

McGillion, M., Pine, J. M., Herbert, J. S., & Matthews, D. (2017). A randomised controlled trial to test the effect of promoting caregiver contingent talk on language development in infants from diverse socioeconomic status backgrounds. *J Child Psychol Psyc, 58*(10), 1122–1131.

Noble, K. G., Engelhardt, L. E., Brito, N. H., Mack, L. J., Nail, E. J., Angal, J., et al. (2015). Socioeconomic disparities in neurocognitive development in the first two years of life. *Dev Psychobiol, 57*(5), 535-551.

Noble, K. G., Houston, S. M., Kan, E., & Sowell, E. R. (2012). Neural correlates of socioeconomic status in the developing human brain. *Dev Sci, 15*(4), 516-527.

Noble, K. G., Wolmetz, M. E., Ochs, L. G., Farah, M. J., & McCandliss, B. D. (2006). Brain-behavior relationships in reading acquisition are modulated by socioeconomic factors. *Dev Sci, 9*(6), 642-654.

Perkins, S. C., Finegood, E. D., & Swain, J. E. (2013). Poverty and language development: Roles of parenting and stress. *Innov Clin Neurosci, 10*(4), 10-19.

Petersen, D. B., & Spencer, T. D. (2012). The narrative language measures: Tools for language screening, progress monitoring, and intervention planning. *Perspect Lang Learn Educ, 19*(4), 119-129.

Raizada, R. D., Richards, T. L., Meltzoff, A., & Kuhl, P. K. (2008). Socioeconomic status predicts hemispheric specialisation of the left inferior frontal gyrus in young children. *Neuroimage, 40*(3), 1392-1401.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Dev Sci, 17*(6), 880-891.

Redcay, E., Haist, F., & Courchesne, E. (2008). Functional neuroimaging of speech perception during a pivotal period in language acquisition. *Dev Sci, 11*(2), 237-252.

Roberts, M. Y., & Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *Am J Speech Lang Pathol, 20*(3), 180-199.

Rodriguez, E. T., & Tamis-LeMonda, C. S. (2011). Trajectories of the home learning environment across the first 5 years: Associations with children's vocabulary and literacy skills at prekindergarten. *Child Dev, 82*(4), 1058-1075.

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Dev, 85*(3), 956-970.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *J Child Lang, 35*(1), 185-205.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Dev, 83*(5), 1762-1774.

Rowe, M. L., Pan, B. A., & Ayoub, C. (2005). Predictors of variation in maternal talk to children: A longitudinal study of low-income families. *Parent Sci Pract, 5*(3), 285-310.

Ryan, J. E. (2013). Poverty as disability and the future of special education law. *Georgetown Law J, 101*(6), 1455-1503.

Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W. T. (2011). Family socioeconomic status and child executive functions: The roles of language, home environment, and single parenthood. *J Int Neuropsychol Soc, 17*(1), 120-132.

Sheridan, M. A., Sarsour, K., Jutte, D., D'Esposito, M., & Boyce, W. T. (2012). The impact of social disparity on prefrontal function in childhood. *PLoS One, 7*(4), e35744.

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: language to infants in structured and naturalistic contexts. *Dev Sci*.

Wechsler, D. (2012). *Wechsler preschool and primary scale of intelligence* (4th ed.). Bloomington, MN: Pearson.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychol Sci, 24*(11), 2143-2152.

Wiig, E. H., Semel, E. M., & Secord, W. (2013). *Clinical evaluation of language fundamentals* (5th ed.). Bloomington, MN: Pearson.

Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., et al. (2009). Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics, 124*(1), 342-349.

Supplementary Information

Method

Sample size justification

Because this is the first study to examine individual relationships between children's language exposure and fMRI measures of language-related brain activation, effect size estimates were not available to inform sample size. However, the behavioral correlations between language input quantity/quality and children's language scores are typically moderate to strong ($0.4 < r < 0.6$) (Hirsh-Pasek et al., 2015; Hoff, 2003; Rowe, 2012; Weisleder & Fernald, 2013). For 80% power to detect such an effect in the expected direction at $\alpha = 0.05$, one would need to recruit 15-36 participants. Similarly, a majority of studies investigating correlations between behavioral measures and fMRI activation using appropriate *independent analyses* report correlations in the 0.5 to 0.7 range, with a median of 0.6 (Vul, Harris, Winkielman, & Pashler, 2009, Figure 5). Given that individual differences analyses (i.e., correlational analyses) have lower power than within-subjects analyses (i.e., condition differences), common sample-size planning tools for fMRI studies are not appropriate for the present power analysis. Instead, power curves specific to Pearson's correlations in the context of fMRI were consulted (Yarkoni & Braver, 2010). By these estimates, one would need to recruit 15-30 participants for the same parameters stated above. Combined, a sample size of 15-36 is recommended to find expected behavioral and neural effects. However, because of the likelihood of publication bias in previously reported effects (e.g., Anderson, Kelley, & Maxwell, 2017), we aimed for the highest end of this range (n = 36).

Statistical Analysis

All statistical analyses (with exception of whole-brain fMRI analyses) were performed in IBM SPSS Statistics version 24. The first approach was to conduct zero-order Pearson's correlations between children's assessment scores, SES demographics, and LENA measures of language exposure. Because all three variables were intercorrelated, we conducted linear regressions to determine which independent variables predicted unique variance in children's language scores, while controlling for the other independent variables. Finally, we conducted bootstrapped mediation with 10,000 iterations using the PROCESS macro for SPSS (Hayes, 2013) to determine whether language exposure mediated the relationship between SES and children's language scores. The same bootstrapping approach was applied to neural activation measures extracted from a region of interest (see main text).

Executive Functioning measure

In addition to the standardized assessments described in the main text, children also completed a non-standardized executive functioning (EF) task. Because EF relies on prefrontal regions adjacent to/overlapping with frontal language regions, EF was included to serve as a covariate/nuisance variable. The Hearts and Flowers version of the dots task (Davidson, Amso, Anderson, & Diamond, 2006) is commonly used to assess EF in both children and adults, because it requires all three EF dimensions (working memory, inhibition, and cognitive flexibility/switching) with simple instructions. Children rested their hands on a handlebar adjusted to finger-distance away from a touch screen computer and completed a practice run of quickly pressing on-screen buttons in this way. Then, a red heart or flower would appear on the right or left side of the screen, and children were instructed to press the button on the *same* side as a heart (congruent condition) and the button on the *opposite* side of a flower (incongruent condition). The task consisted of three consecutive blocks: a congruent block of 12 trials, an incongruent block of 12 trials, and a randomly mixed block (congruent and incongruent) of 49 trials. For all conditions, stimuli were displayed for 500 milliseconds (ms) with 1500 ms to respond and an interstimulus interval of 500 ms. Any response faster than 200 ms were considered to be anticipatory (Davidson et al., 2006) and excluded from analyses. Children received up to 12 practice trials before the congruent and incongruent blocks to ensure understanding of the rule. No practice was included before the mixed block, and thus the first trial was additionally excluded from analyses. The main outcome measures were the average accuracy across all trials in the mixed block and average reaction time (RT) in milliseconds across all correctly answered trials in the mixed block. Although rare, accuracy scores below 50% were not discarded because they could have been obtained by rule reversal, indicating an error in cognitive flexibility/switching.

Results

Mean accuracy on the EF task was 72%, (*SD* = 22%) with a mean reaction time on correctly answered trials of 1140 ms (*SD* = 207). None of the fMRI analyses – including group mean task activation and whole brain correlates with LENA measures – changed with the inclusion of EF scores as a nuisance variable. This suggests that correlations between conversational turns and activation in Broca's area are not driven by differences in executive functioning.

Supplementary References

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychol Sci*, 956797617723724.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037-2078.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: The Guilford Press.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff , R. M., Pace, A., et al. (2015). The contribution of early communication quality to low-income children's language success. *Psychol Sci, 26*(7), 1071-1083.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Dev, 74*(5), 1368-1378.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Dev, 83*(5), 1762-1774.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect Psychol Sci, 4*(3), 274-290.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychol Sci, 24*(11), 2143-2152.

Yarkoni, T., & Braver, T. S. (2010). Cognitive Neuroscience Approaches to Individual Differences in Working Memory and Executive Control: Conceptual and Methodological Issues. In A. Gruszka, G. Matthews & B. Szymura (Eds.), *Handbook of Individual Differences in Cognition: Attention, Memory, and Executive Control* (pp. 87-107). New York, NY: Springer New York.

Table S1. Group mean task activations for forward > backward speech

	MNI coordinates			
Z-value	x	y	z	Anatomical Description
Left Hemisphere Cluster (3552 voxels)				
6.40	-52	-7	-12	Anterior superior temporal sulcus
6.39	-55	-39	4	Posterior superior temporal sulcus
6.11	-50	9	-22	Temporal pole
5.81	-57	-51	30	Supramarginal gyrus
Right Hemisphere Cluster (1418 voxels)				
5.93	54	-4	-13	Anterior superior temporal sulcus
5.47	51	13	-20	Temporal pole

Coordinates and anatomical descriptions of local peak activations for the forward > backward speech contrast, averaged over the entire sample (n = 36). Analyses revealed two significant clusters, one in each hemisphere, visualized in Figure 4.

Table S2. Correlation between conversational turns and forward > backward activation

	MNI coordinates			
Z-value	x	y	z	Anatomical Description
4.59	-48	33	15	Left posterior pars triangularis
4.18	-56	33	10	Left anterior pars triangularis
3.55	-43	13	15	Left anterior pars opercularis

Coordinates and anatomical descriptions of local peak activations in a single cluster (Figure 5a, 766 voxels) exhibiting a significant correlation between the number of conversational turns children experienced per hour and activation in the forward > backward speech contrast.