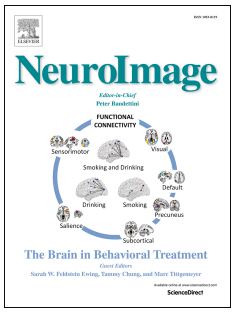
Journal Pre-proof



A meta-analysis of fMRI studies of language comprehension in children

Alexander Enge, Angela D. Friederici, Michael A. Skeide

PII: S1053-8119(20)30344-X

DOI: https://doi.org/10.1016/j.neuroimage.2020.116858

Reference: YNIMG 116858

To appear in: *NeuroImage*

Received Date: 24 January 2020

Revised Date: 10 April 2020

Accepted Date: 11 April 2020

Please cite this article as: Enge, A., Friederici, A.D., Skeide, M.A., A meta-analysis of fMRI studies of language comprehension in children, *NeuroImage*, https://doi.org/10.1016/j.neuroimage.2020.116858.

A meta-analysis of fMRI studies of language comprehension in children

Alexander Enge, Angela D. Friederici, Michael A. Skeide

Corresponding author:

Michael A. Skeide, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1A, 04103 Leipzig, Germany, +49 (0)341 9940 130, skeide@cbs.mpg.de

CRediT author statement

Alexander Enge: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing - Review & Editing, Visualization

Angela D. Friederici: Conceptualization, Writing - Review & Editing

Michael A. Skeide: Conceptualization, Methodology, Investigation, Writing - Original Draft, Supervision, Project Administration

A meta-analysis of fMRI studies of language comprehension in children

Running title: Language fMRI meta-analysis in children

Alexander Enge,[1,2] Angela D. Friederici,[1] Michael A. Skeide[1,*]

[1]Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany

[2]Department of Psychology, Humboldt-Universität zu Berlin, 12489 Berlin, Germany

[*]Corresponding author: Michael A. Skeide, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1A, 04103 Leipzig, Germany, +49 (0)341 9940 130, skeide@cbs.mpg.de

Abstract

The neural representation of language comprehension has been examined in several meta-analyses of fMRI studies with human adults. To complement this work from a developmental perspective, we conducted a meta-analysis of fMRI studies of auditory language comprehension in human children. Our analysis included 27 independent experiments involving n = 625 children (49% girls) with a mean age of 8.9 years. Activation likelihood estimation and seed-based effect size mapping revealed activation peaks in the pars triangularis of the left inferior frontal gyrus and bilateral superior and middle temporal gyri. In contrast to this distribution of activation in children, previous work in adults found activation peaks in the pars opercularis of the left inferior frontal gyrus and more left-lateralized temporal activation peaks. Accordingly, brain responses during language comprehension may shift from bilateral temporal and left pars triangularis peaks in childhood to left temporal and pars opercularis peaks in adulthood. This shift could be related to the gradually increasing sensitivity of the developing brain to syntactic information.

Introduction

Hemodynamic activity during language comprehension has been extensively examined using fMRI in adults. Since typical sample sizes of single studies range from about 10 to 30 participants, meta-analytic methods have been used to increase statistical power and detect robust effects across experiments (Binder et al., 2009; Ferstl et al., 2008; Rodd et al., 2015; Vigneau et al., 2006). Following this approach, three canonical regions underlying language comprehension in adults were consistently found: the left inferior frontal gyrus (IFG), the left middle temporal gyrus (MTG), and the left superior temporal gyrus (STG; Binder et al., 2009; Ferstl et al., 2008; Rodd et al., 2015; Vigneau et al., 2006). While the pars opercularis of the left IFG and the left STG was related to syntactic processing (Rodd et al., 2015; Vigneau et al., 2006), the pars triangularis and orbitalis of the left IFG and the left MTG was related to semantic processing (Binder et al., 2009; Ferstl et al., 2008). When pooling across tasks, peak activation was localized in the pars opercularis of the left IFG and MTG (Rodd et al., 2015).

A growing body of literature has reported fMRI results obtained from language comprehension experiments with children. At the word level, experiments have targeted phonological processing with the two-word rhyme judgement task (Cao et al., 2008; Cone et al., 2008; Desroches et al., 2010) and the first-sound matching task (Raschle et al., 2014), semantic processing with the noun categorization task (Balsamo et al., 2006), and (morpho-)syntactic processing with the morphological awareness task (Arredondo et al., 2015). At the sentence level, the description definition task has been frequently used (Bartha-Doering et al., 2018; Berl et al., 2014; Moore-Parks et al., 2010). In this task, children listen to a noun preceded by a short description. They are asked to judge whether the two are matching (e.g. “A long yellow fruit is a banana”) or not (e.g. “Something you sit on is a spaghetti”). Other tasks include a judgment on whether two semantically or syntactically manipulated sentences convey the same meaning (Borofsky et al., 2010; Nuñez et al., 2011), the semantic/syntactic acceptability task (Brauer et al., 2011), and the sentence-picture matching task (Horowitz-Kraus et al., 2015). Moreover, the passive listening task has also been employed in a number of studies (Knoll et al., 2012; Monzalvo et al., 2012). This task has also often been used for stories (Horowitz-Kraus et al., 2016; Romeo et al., 2018; Sroka et al., 2015),

sometimes additionally asking children to answer content-related questions or retell the stories (Vannest et al., 2019).

While meta-analyses of fMRI studies of language comprehension in children are currently not available, the present literature is synthesized in a qualitative and a quantitative review (Skeide & Friederici, 2016; Weiss-Croft & Baldeweg, 2015). These reviews suggest that activity in left-lateralized regions in the IFG, MTG, and STG known from the adult literature is already broadly established by 3 years of age. Focused activity in the pars opercularis of the left IFG, however, emerges only gradually towards adulthood when children become more sensitive to syntactic information (e.g. morphology, word order). Younger children, in contrast, rely more on semantic information and thus more strongly recruit the pars triangularis of the left IFG (Skeide & Friederici, 2016; Weiss-Croft & Baldeweg, 2015).

Here we conducted the first statistical synthesis of the fMRI literature on language comprehension in children. To this end, we quantified the overlap of hemodynamic activations reported in previous studies using activation likelihood estimation (ALE; Eickhoff et al., 2012; Turkeltaub et al., 2002) and seed-based effect size mapping (SDM; Albajes-Eizagirre et al., 2019). Following the currently available original and review articles, we hypothesized two major differences in the activation patterns associated with language processing in children compared to adults. First, we expected activation peaks in the temporal cortex to be less strongly lateralized to the left hemisphere. This hypothesis was based on a number of individual studies in children which have found significant clusters of activation not only in the left MTG and STG, but also in the right MTG and STG. The latter effect is typically not consistently found in adults (e.g. Holland et al., 2007; Horowitz-Kraus et al., 2015; Sroka et al., 2015; Szaflarski et al., 2006). Second, we hypothesized that the distribution of activity in the left IFG would differ such that peak activity in children would be found in the pars triangularis, while peak activity in adults would be found in the pars opercularis. This hypothesis was based on previous work indicating that at least until their first years in school, children rely more strongly on semantic information—typically associated with enhanced recruitment of the pars triangularis of the left IFG—and only gradually become more sensitive to

64 syntactic information—typically associated with peaks of activation in the pars opercularis of the left IFG
65 (e.g. Nuñez et al., 2011; Skeide et al., 2014).

66 **Materials and Methods**

67 ***Literature Search***

68 The PubMed database (https://www.ncbi.nlm.nih.gov/pubmed/) was used to identify articles
69 containing the terms "fMRI AND language AND children" or "functional MRI AND language AND
70 children" in their respective title or abstract. As of August 2019, this search yielded 356 results after
71 removing duplicate entries. These results were screened to exclude any articles that did not meet one or
72 more of the following predefined inclusion criteria:

73 (1) The article was written in English.
74 (2) There was at least one group of healthy, monolingual children with a mean age between 3
75 and 15 years. The lower boundary of this age range was set by the feasibility of task-based
76 fMRI studies in children and the upper boundary was chosen to include a gap of 3 years until
77 adulthood (18 years).
78 (3) The children completed a natural language task during fMRI scanning, not an artificial
79 language task (e.g. scrambled syllables).
80 (4) The authors conducted a random-effects analysis using a general linear model to obtain
81 whole-brain within-group results.
82 (5) Peak coordinates were reported in Tailarach or Montreal Neurological Institute (MNI) space.

83 Thus, articles only reporting the results of conjunction analyses of multiple tasks (e.g. reading and
84 listening tasks) or groups (e.g. children and adults or children with and without reading difficulty) were
85 excluded to maintain the main focus of our analysis on the activation associated with specific types of
86 language tasks in typically developing children. Applying these criteria, we identified 41 eligible articles,
87 37 of which used language comprehension tasks, whereas four deployed language production tasks (e.g.
88 overt or covert verb generation tasks). As previously noted by Weiss-Croft & Baldeweg (2015), the small

number of language production experiments might be explained by the problem of speech-related movement artifacts, which is aggravated in studies with children.

In a recent simulation study, Eickhoff et al. (2016) demonstrated that at least 17 to 20 independent experiments should be included in any ALE-based meta-analysis of neuroimaging data. This ensures acceptable robustness and power by minimizing the chance that the results are driven by single experimental results and by maximizing the chance to detect small- and medium-size effects. Accordingly, we were not able to run separate analyses for both comprehension and production tasks. Instead, we decided to narrow down our analysis to those articles investigating language comprehension. Of these, 27 articles included at least one condition in which stimuli (words, sentences, or stories) were presented auditorily, whereas ten articles exclusively used visual stimuli. We excluded these ten reading experiments because, as before, their number was insufficient to conduct a robust and sufficiently powered ALE-based meta-analyses (Eickhoff et al., 2016), thus focusing our analysis on auditory language comprehension experiments. Finally, two of the remaining articles were excluded because fMRI was recorded while children were asleep in the scanner (Redcay et al., 2008). One further article (Monzalvo et al., 2012) was excluded because the same contrast and group of children had already been included as part of another, more comprehensive publication (Monzalvo and Dehaene-Lambertz, 2013). This entire selection process, which is summarized as a flowchart in Figure 1, yielded a final sample of 24 articles that could be included in the present meta-analysis.

Post hoc, we included various related search terms (“brain”, “function*”, “magnet*”, “BOLD”, “child”, “development”, and all names of the tasks used in the already included studies). Furthermore, we screened the reference lists and tables of the original articles and reviews. These screening procedures did not reveal any additional suitable studies that were not yet captured by the initial PubMed search.

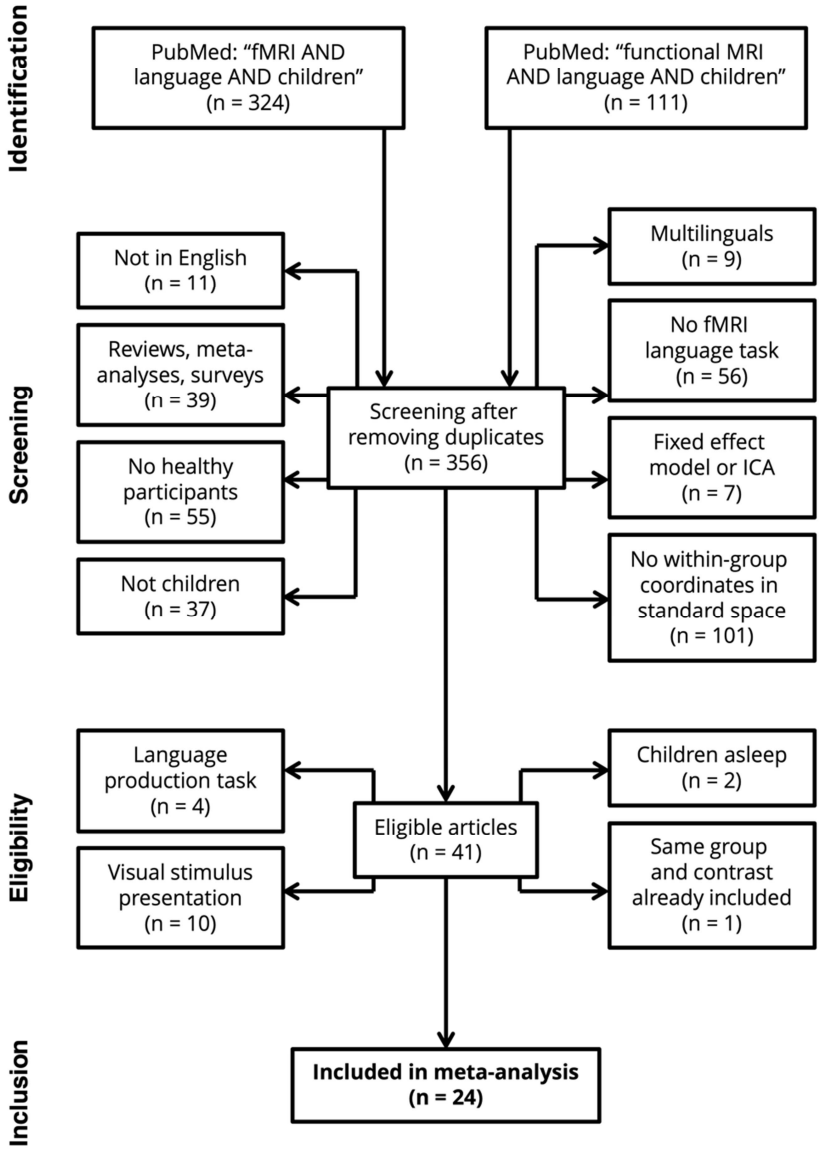


Figure 1. PRISMA flowchart of the selection process for included articles. An initial screening of the 356 articles listed on PubMed (as of August 2019) revealed 41 articles that reported whole-brain coordinates in standard space obtained from fMRI experiments targeting language processing in healthy children. Of these, 24 articles using auditory language comprehension tasks were included in the statistical analysis. ICA = independent component analysis.

Activation Likelihood Estimation

To identify converging activation across the experiments reported in these articles, we conducted an activation likelihood estimation (ALE; Eickhoff et al., 2012; Turkeltaub et al., 2002) as implemented

in the GingerALE software, version 3.0.2 (http://brainmap.org/ale/). ALE performs a coordinate-based meta-analysis of the peak coordinates reported in fMRI experiments to determine where in the brain results converge at an above-chance level. "Experiment" refers to one type of task (i.e. auditory language comprehension) in one specific sample (i.e. one group of children). Hence, multiple fMRI contrasts reported within a single article constitute multiple independent experiments if they are obtained from different samples, but should be pooled into a single experiment when they are obtained from the same sample to control for within-group effects (Eickhoff et al., 2017; Turkeltaub et al., 2012). In our meta-analysis, the 24 articles reported 32 contrasts for auditory language comprehension. Ten of these contrasts were investigated in identical or overlapping samples and their foci were thus pooled into one experiment (see Table 1). This procedure resulted in 27 experiments reporting 453 foci in total (mean = 16.8, median = 11 per experiment). Eight of these experiments reported foci in Talairach space and were converted to MNI space using the *icbm2cal* function as implemented in GingerALE (Lancaster et al., 2007).

As a first step, the ALE algorithm created a binary map for each experiment in which all activated voxels were assigned a value of 1 and all other voxels are assigned a value of 0. Next, to account for the uncertainty associated with using condensed peak information instead of parametric whole-brain maps, a three-dimensional Gaussian distribution was fitted around each of these peaks, smoothing out their activation across the neighboring voxels. ALE determines the amount of uncertainty based on the sample size of the respective experiment, with foci from larger samples being smoothed with a narrower kernel (Eickhoff et al., 2009). This resulted in separate mean activation maps for all experiments, which were then combined into a single ALE map using a random-effects approach. To do so, each voxel was assigned a value corresponding to the union of its activation probabilities from the individual mean activation maps. This so-called ALE value indicates, for each gray matter voxel, the degree of convergence in activation between all included experiments. Finally, the map of ALE values was statistically thresholded to check in which voxels convergence could be expected to be above chance level. As recommended on the basis of a recent simulation by Eickhoff et al. (2016), we combined an uncorrected cluster-forming voxel-wise height threshold of $p < .001$ and a cluster-wise family-wise error

(FWE) correction with a threshold of $p < .05$ based on 1,000 random permutations. All voxels surviving this threshold were interpreted as showing above-chance convergence between experiments reflecting the "true" activation associated with auditory language comprehension in children. Local peaks within these significant clusters were assigned their respective anatomical gray matter labels using the Anatomy toolbox, version 2.2c (Eickhoff et al., 2007, 2006, 2005) in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/). This toolbox provides anatomical labels for peak coordinates in MNI space based on probabilistic maps and, for peaks within the IFG, their probability (in %) of belonging to Brodmann Area 44 (pars opercularis) or 45 (pars triangularis; Amunts et al., 2004). Finally, the Talairach Daemon atlas as implemented in GingerALE (Lancaster et al., 2000, 1997) was used to determine the Brodmann Area of peaks outside the IFG.

Comparison with Previous Adult Meta-Analysis

The pattern of activation associated with language comprehension in children was compared to previous meta-analytic work on language comprehension in adults. To achieve this, we reproduced the meta-analysis by Rodd et al. (2015), which included 54 studies on semantic and syntactic language processing with a total of 957 adult subjects and 320 foci. Details of the literature search, inclusion criteria, and meta-analytic methods can be found in the original publication (Rodd et al., 2015). For the purpose of the present study, we deviated from the original analysis in two aspects. First, we excluded any experiments using visual stimuli (i.e. reading experiments), in line with our meta-analysis in children, which included only experiments using auditory stimuli. This resulted in a subset of 23 studies in adults with a total of 431 subjects and 105 foci. Second, in the original publication, data were thresholded and corrected based on the false discovery rate. In contrast, here we used an uncorrected cluster-forming voxel-wise height threshold of $p < .001$ and a cluster-wise FWE-corrected threshold of $p < .05$, which is identical to the threshold that we used for analyzing the data of the children. The cluster-wise FWE-corrected threshold was preferred in accordance with a recent simulation study by Eickhoff et al. (2016). These authors demonstrated that this threshold provides the highest statistical power and therefore the

highest sensitivity to detect "true" effects that were known a priori by simulating the data. At the same time, this threshold turned out not to inflate the number of spuriously significant clusters. Thresholding using the false discovery rate as in Rodd et al. (2015), on the other hand, was shown to lead to both substantially reduced statistical power and an increase in the number of spurious clusters.

After obtaining the thresholded ALE map of the adult experiments, we compared it statistically to the ALE map of experiments in children. To this end, we subtracted ALE maps to identify clusters where activation was found more consistently in one group compared to the other group (children > adults, adults > children). Additionally, we created a conjunction map showing similarities in activation between the two groups. In each case, the resulting ALE map was thresholded using an uncorrected cluster-forming voxel-wise height threshold of $p < .001$ and a cluster-wise FWE-corrected threshold of $p < .05$. Clusters that were significant at this level were anatomically labeled using the Anatomy toolbox in SPM12 and assigned to a Brodmann Area based on the Talairach Daemon atlas.

Seed-Based Effect Size Mapping

An alternative approach to statistically synthesize results from multiple fMRI experiments is seed-based effect size mapping (SDM; Albajes-Eizagirre et al., 2019). Similar to ALE, SDM uses a coordinate-based random-effects approach that combines the information of peak coordinates in standard space across multiple experiments. While ALE treats all peak coordinates the same, SDM accounts for the effect size associated with each peak and reconstructs the original parametric maps of the individual experiments before combining them into a meta-analytic map. Hence, while ALE maps quantify the degree of overlap in peak activation across experiments, SDM estimates the effect size of activation or deactivation for each voxel. Although the SDM method is still less commonly used compared to ALE (Acar et al., 2018), we thought it might complement our main results in three aspects. First, the fact that SDM uses a different algorithm than ALE renders it possible to scrutinize the robustness and replicability of the results obtained from ALE. Second, SDM differentiates between voxels with significant activation and deactivation while ALE only captures activation (Radua & Mataix-Cols, 2012). Finally, SDM makes

it possible to include covariates and compute meta-regression analyses as a means to estimate the influence of potentially confounding variables.

We performed the additional SDM meta-analysis using the same peak coordinates as before but adding, whenever possible, their associated *t*- or *z*-values (the latter being converted to a *t*-value). This analysis was conducted using the SDM-PSI software, version 6.11 (https://www.sdmproject.com/). First, effect size maps were built for the 27 individual experiments. This was accomplished by (a) converting the *t*-value of each peak coordinate into an estimate of effect size (Hedge's *g*) using standard formulas (Hedges, 1981) and (b) convolving these peaks with a fully anisotropic unnormalized Gaussian kernel ($\alpha = 1$, FWHM = 20 mm) within the boundaries of the default gray matter template as provided by SDM (voxel size = 2×2×2 mm). Effect sizes for peaks with unknown *t*- or *z*-values were estimated from a threshold-based imputation based on the mean effect size of peaks for which *t*-values are known. Imputation was conducted separately for groups of experiments with different statistical thresholds (Radua et al., 2012). Second, the individual effect size maps were combined using a random-effects general linear model. Third, the statistical significance of activations in the resulting meta-analytic effect size map was examined by comparing it to 1,000 random permutations of activation peaks within the gray matter template. Finally, the meta-analytic maps were thresholded using an uncorrected voxel-wise height threshold of $p < 0.001$ and a cluster-wise extent threshold of $k = 50$ voxels, which approximately corresponds to the FWE-corrected thresholding procedure implemented in ALE (Eickhoff et al., 2012; Radua et al., 2012). Peak coordinates of the resulting meta-analytic clusters of activation were anatomically labeled using the Anatomy toolbox in SPM12 and assigned to a Brodmann Area based on the Talairach Daemon atlas.

SDM was also used to assess the effect of four potentially confounding variables on the results of the meta-analysis, namely, age (mean age of children in each experiment), baseline (1 = rest/fixation, 2 = active), type of language task (1 = story listening, 2 = decision tasks at the sentence level, 3 = decision tasks at the word level), and software package used for image processing and statistical analysis in the original publication (1 = SPM, 2 = FSL/LIPSIA/AFNI). For each the four variables, a separate linear

model was calculated in SDM to identify clusters that significantly covaried with the respective variable. All pre-processing and thresholding parameters were kept the same as in the main analysis.

Jackknife Sensitivity Analysis

To explore how potentially spurious results in the literature (e.g. driven by publication bias) would affect the results of our ALE analysis, we conducted a Jackknife sensitivity analysis. To this end, we ran 27 different meta-analyses in ALE, each with a different experiment of the original sample being left out. We visually inspected how well each of these simulations reproduced the original results in terms of number, location, and size of significant ALE voxels. Substantial variability would indicate that the results are driven by the specific study that had been left out, thus compromising the robustness to spurious (e.g. false positive or *p*-hacked) findings.

Fail-Safe N Analysis

To further evaluate the robustness of the present results against unpublished studies with null results in the "file drawer" (e.g. driven by bias towards publishing positive results), we carried out a fail-safe N analysis. The rationale behind this approach is to investigate the effect of iteratively adding null-result experiments to our original sample (Acar et al., 2018). Null-result experiments were created in R, version 3.6.1 (https://www.r-project.org), matching the real experiments in terms of sample size and number of foci reported, but with foci being distributed randomly across the gray matter. Next, new meta-analyses were computed in ALE by iteratively adding one null experiment after another to the original data. For each significant cluster in the original analysis, the fail-safe N was defined as the highest number of null experiments that could be added until the cluster failed to reach statistical significance. Thus, fail-safe N indicates how many fMRI studies with non-significant results could be hidden in the file drawer without compromising the significance of a certain cluster. To increase reliability, the whole procedure was repeated with ten different, randomly generated sets of null-result experiments, each

representing one potential file drawer. The mean fail-safe N of these ten simulations was calculated separately for each cluster.

Following Acar et al. (2018), we also pre-specified lower and upper boundaries for the fail-safe N of each cluster based on the following considerations. A recent modelling approach to data from the BrainMap database (http://brainmap.org/) indicates that there might be up to 30 unpublished null studies per 100 published neuroimaging studies in the language domain (Samartsidis et al., 2019). Using this conservative estimate of the file drawer effect, we pre-specified that the fail-safe N for each cluster should exceed a lower boundary of eight added null experiments (equaling 30% of the real data). The upper boundary was pre-specified as the number of null experiments that could be added so that the real experiments still made up for 10% or more of the foci contributing to a particular cluster. This ensures that the significance of a cluster is driven by the majority of experiments instead of few highly influential ones. Only if the actual fail-safe N obtained from the simulation is between these two boundaries, the cluster can be assumed to be robust against both a potential file drawer effect and hyper-influential effects of a few experiments.

Results

Descriptive Statistics

Twenty-seven experiments reported in 24 articles published between 2003 and 2019 were included in the present meta-analysis. Participants were 625 typically developing, monolingual children with a mean age of 8.9 years (range: 3 to 15 years). Gender was approximately equally distributed (49% females) and children were almost exclusively right-handed (96%). Of the 27 experiments, eight involved judgments at the word level, 12 involved judgments at the sentence level, and seven involved listening to spoken stories. A descriptive overview of these experiments is provided in Table 1 and the distributions of mean ages and sample sizes of the experiments are depicted in Figure 2.

Table 1

Descriptive Information of the 27 Experiments Included in the Meta-Analysis

No.	Article	Sample[a] (*n*)	Age (mean, range)	Females (*n*)	Handed-ness[c]	Task	Foci (*n*)
1	Ahmad et al. (2003)	15	6.8, 5–7	9	R	Story > reversed story	3
2	Arredondo et al. (2015)	16	9.3, 6–12	8	R	Word repetition judgment > rest	10[d]
						Morphological judgment > rest	
3	Balsamo et al. (2006)	23	8.5, 5–10	13	1L	Noun categorization > reversed nouns	9
4	Bartha-Doering et al. (2018)	30	10.3, n/r[b]	12	R	Description-definition matching > reversed speech	7
5	Bartha-Doering et al. (2019)	18	10.9, n/r[b]	7	R	Description-definition matching > reversed speech	11
6	Berl et al. (2010)	44	10, 7–12	20	R	Story > reversed story	6
7	Berl et al. (2014)	57	8.9, 4–12	29	R	Description-definition matching > reversed speech	15
8	Borofsky et al. (2010)	14	12.4, n/r[b]	8	R	Two-sentence matching > rest	63[d]
						Two-sentence matching > rest	
9	Brauer et al. (2011)	10	7, 5–8	5	R	Sentence acceptability judgment > rest	39
10	Brennan et al. (2013)	16	10.3, 8–12	n/r[b]	R	Rhyme judgment > fixation cross	22
11		15	10.3, 8–12	n/r[b]	R	Rhyme judgment > fixation cross	30
12	Cao et al. (2011)	25	10.4, 8–12	12	R	Rhyme judgment > tone judgment	20[d]
						Spelling judgment > tone judgment	
13	Cone et al. (2008)	40	11.9, 9–15	22	R	Rhyme judgment > tone judgment	8
14	Desroches et al. (2010)	12	11.5, 8–14	4	R	Rhyme judgment > fixation cross	9
15	Horowitz-Kraus et al. (2015)	23	8.5, n/r[b]	15	R	Sentence-picture matching > word-picture matching	28
16	Horowitz-Kraus et al. (2016)	9	10.2, n/r[b]	7	2L	Story listening > broadband noise sweep listening	9
17	Hubbard et al. (2012)	10	12.1, n/r[b]	0	R	Story listening + picture viewing > picture viewing	8
18	Knoll et al. (2012)	22	5.8, 4–6	9	6A	Sentence listening subject initial) > rest	22[d]
						Sentence listening object initial) > rest	
19	Monzalvo & Dehaene-Lambertz (2013)	23	9.6, 8–10	11	2L	Sentence listening > Foreign sentence listening	9
20		13	6.8, 5–7	6	2L	Sentence listening > Foreign sentence listening	11
21		13	6.2, 5–6	7	3L	Sentence listening > Foreign sentence listening	8
22	Moore-Parks et al. (2010)	23	8.8, 7–10	12	R	Description-definition matching > reversed speech	15
23	Nuñez et al. (2011)	19	11.1, 7–15	10	1L	Two-sentence matching > rest	27
24	Raschle et al. (2014)	20	5.9, 5–6	n/r[b]	1A	Voice matching > rest	34[d]
						First-sound matching > rest	
25	Romeo et al. (2018)	36	5.8, 4–6	22	n/r[b]	Story listening > reversed speech	6
26	Sroka et al. (2015)	30	4.2, 3–5	17	n/r[b]	Story listening > broadband noise sweep listening	9
27	Vannest et al. (2019)	40	7.9, 5–12	20	R	Story listening > broadband noise sweep listening	15

[a]only typically developing children [b]not reported [c]R = right-handed, L = left-handed, A = ambidextrous, [d]contrasts of the same article were treated as a single experiment due to identical or overlapping samples of children

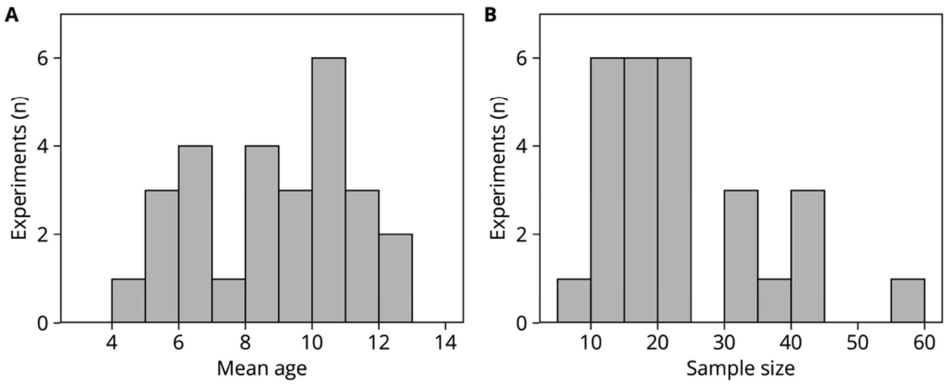


Figure 2. Distributions of the mean ages (A) and sample sizes (B) of the 27 experiments included in the meta-analysis.

Activation Likelihood Estimation

Five activation clusters associated with auditory language comprehension in children showed significant convergence across the experiments ($p < .05$, cluster-wise FWE-corrected). The largest peak was found in the pars triangularis of the left IFG (Brodmann Area [BA] 45). The corresponding cluster extended across the pars opercularis of the left IFG (BA 44) to left middle and superior frontal cortices (BA 46, BA 6), and left precentral cortices (BA 6). Moreover, the cluster extended across the pars orbitalis of the left IFG (BA 47) to the left insula (BA 13). A smaller cluster was detected in the pars triangularis and the pars orbitalis of the right IFG and the right insula. Two other clusters covered the STG (BA 22, BA 41) and the MTG (BA 21, BA 38) bilaterally. Finally, one more cluster was identified in left premotor and anterior cingulate regions (BA6, BA8, BA32, BA24; Figure 3, Table 2).

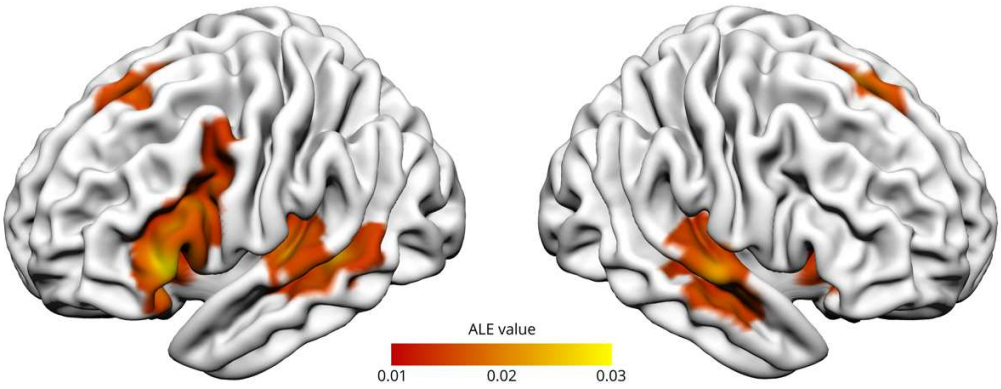


Figure 3. ALE map of significant clusters associated with language comprehension in children, superimposed onto a standard cortical surface. Activations reported in 27 experiments that showed above-chance overlap ($p < .05$, cluster-wise family-wise error [FWE] corrected) are shown. The color bar represents the ALE value of any given voxel, that is, its degree of non-random convergence in activation between experiments.

Table 2

Local Peaks and Descriptive Information of the Five Clusters with Above-Chance Overlap

Cluster	Size (mm^3)	ALE (peak)	*z* (peak)	x	y	z	Anatomical location[a] (probability)[b]	BA[c]	Contributing experiments[d]
1	8,328	0.037	6.30	-52	28	8	Left Inferior Frontal Gyrus, pars triangularis (BA45-56%)	45	2|3|4|5|6|7|8|9|12|13|15|16 |18|20|21|23|24|27
		0.028	5.16	-42	28	-4	Left Inferior Frontal Gyrus, pars orbitalis	47	
		0.028	5.16	-48	24	16	Left Inferior Frontal Gyrus, pars triangularis (BA45-29%)	46	
		0.025	4.82	-44	14	22	Left Inferior Frontal Gyrus, pars opercularis (BA44-18%, BA45-1%)	9	
		0.022	4.34	-42	6	30	Left Precentral Gyrus (BA44-16%)	6	
		0.019	3.95	-28	26	0	Left Insula Lobe	-[e]	
		0.017	3.55	-42	4	42	Left Precentral Gyrus	6	
		0.015	3.24	-44	2	50	Left Precentral Gyrus	6	
2	4,872	0.031	5.54	-52	-38	2	Left Middle Temporal Gyrus	22	1|2|4|6|8|11|12|13|14|17|18 |19|20|21|22|24|25|27
		0.030	5.44	-56	-22	2	Left Superior Temporal Gyrus	41	
3	4,856	0.028	5.19	64	-12	0	Right Superior Temporal Gyrus	22	2|7|8|11|12|13|14|16|17|18 |19|20|21|22|24|25|27
		0.022	4.40	56	-16	-8	Right Superior Temporal Gyrus	22	
		0.022	4.35	54	-6	-12	Right Superior Temporal Gyrus	22	
		0.020	4.04	54	-26	4	Right Superior Temporal Gyrus	41	
		0.020	4.01	46	-24	8	Right Heschl's Gyrus	13	
4	3,992	0.050	7.76	-4	14	52	Left Posterior Medial Frontal Gyrus	6	2|3|5|7|8|10|11|13|14|15| 18|22|23|24
		0.017	3.60	2	30	42	Left Superior Medial Frontal Gyrus	8	
5	1,808	0.026	4.95	32	30	-2	Right Insula Lobe (BA45-9%)	45	4|7|8|9|10|12|18|22|24
		0.020	4.11	38	24	4	No anatomical label found	13	

[a]according to the SPM Anatomy toolbox [b]probability of belonging to Brodmann area (BA) 44 or 45 [c]Brodmann area according to Talairach Daemon [d]see Table 1 [e]no Brodmann area found by Talairach Daemon

Comparison with Adult Meta-Analysis

A comparison of this pattern of activations associated with language comprehension in children to the pattern observed in adults revealed a number of similarities, including clusters of common activation in the left IFG (BA 13, BA 45), the left MTG and STG (BA 22), the right STG (BA 13, BA 22, BA 41), and the left medial frontal gyrus (BA 6, BA 9; Figures 4 and 5, Table 3).

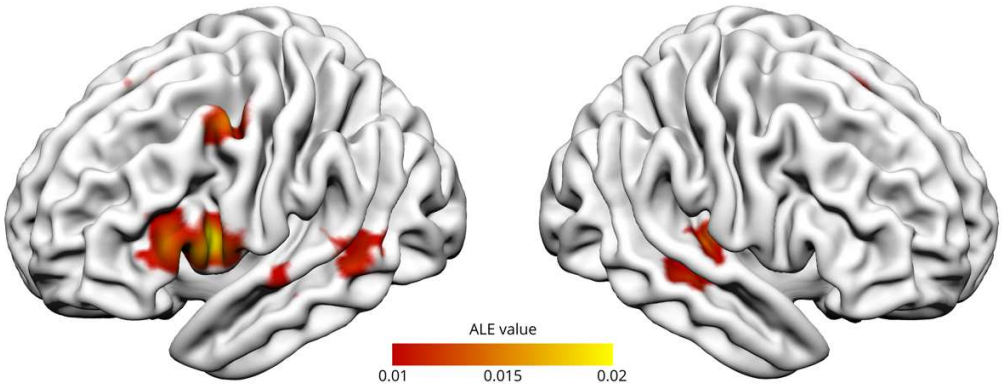


Figure 4. ALE map of significant clusters associated with language comprehension in adults. These data were reproduced using the sample of studies reported in a previous meta-analysis by Rodd et al. (2015). Maps depict clusters with above-chance overlap ($p < .05$, cluster-wise FWE-corrected) and their associated ALE value (color bar), that is, the degree of non-random convergence in activation between experiments at any given voxel.

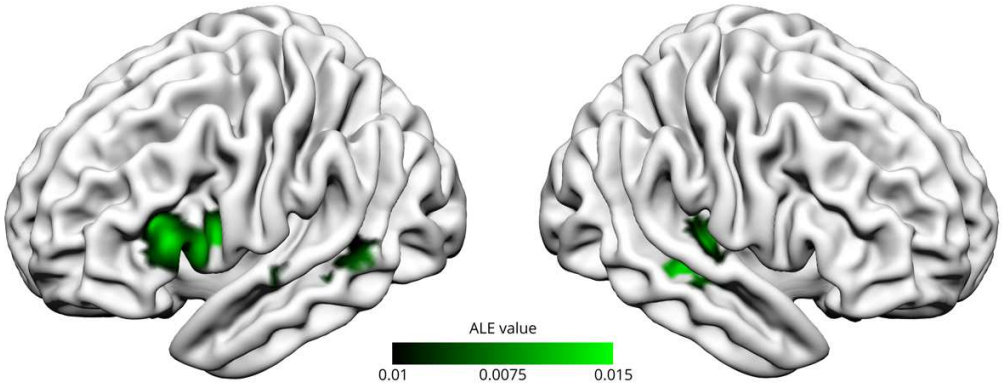


Figure 5. ALE map of significant clusters associated with language comprehension in both children and adults as a result of a conjunction analysis. Maps depict clusters with above-chance overlap ($p < .05$, cluster-wise FWE-corrected) in the ALE maps of both children (Figure 3) and adults (Figure 4). The color bar represents the voxel-wise minimum convergence between these two images.

Table 3

Group Similarities Between Children and Adults

Conjunction

Cluster	Size (mm^3)	ALE (peak)	x	y	z	Anatomical location[a] (probability)[b]	BA[c]
1	2,176	0.021	-48	26	14	Left Inferior Frontal Gyrus, pars triangularis (BA45-35%)	45
		0.019	-54	22	10	Left Inferior Frontal Gyrus, pars triangularis (BA45-38%, BA44-9%)	45
		0.018	-46	14	16	Left Inferior Frontal Gyrus, pars opercularis (BA44-34%, BA45-2%)	13
2	832	0.018	46	-24	6	Right Superior Temporal Gyrus	13
		0.016	54	-26	2	Right Superior Temporal Gyrus	41
		0.015	56	-18	0	Right Superior Temporal Gyrus	22
		0.015	54	-22	2	Right Superior Temporal Gyrus	41
3	632	0.016	-2	14	54	Left Posterior-Medial Frontal Gyrus	6
4	272	0.015	-58	-40	6	Left Middle Temporal Gyrus	22
		0.012	-54	-46	4	Left Middle Temporal Gyrus	22
5	248	0.013	-56	-28	2	Left Middle Temporal Gyrus	22
6	120	0.011	-60	-14	0	Left Middle Temporal Gyrus	22
		0.009	-54	-16	2	Left Superior Temporal Gyrus	22
7	24	0.010	-48	-22	0	Left Middle Temporal Gyrus	-[d]
8	8	0.009	-40	18	22	Left Inferior Frontal Gyrus, pars triangularis (BA44-6%, BA45-5%)	9

[a]according to the SPM Anatomy toolbox [b]probability of belonging to Brodmann area (BA) 44 or 45 [c]Brodmann area according to Talairach Daemon [d]no Brodmann area found by Talairach Daemon

Children revealed significantly more consistent activation in the right STG and MTG (BA 21, BA 22), the left medial and superior frontal gyri (BA 6, BA 8, BA 9), the pars triangularis of the IFG (BA 45), the left STG and MTG (BA 21, BA 41), and the left and right insulae (BA 13). Adults showed more consistent activation than children in the pars opercularis of the left IFG (BA 44; Figure 6, Table 4).

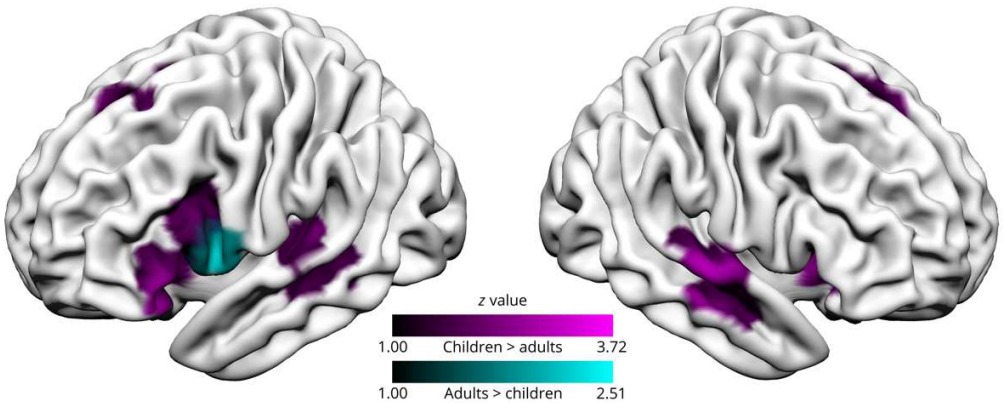


Figure 6. Contrast map of regions where activation associated with auditory language comprehension was more consistent in children than in adults (purple) or more consistent in adults than in children (cyan). Individual ALE maps (Figures 3 and 4) were subtracted from one another and thresholded at $p < .05$ (cluster-wise FWE-corrected).

Table 4

Group Differences Between Children and Adults

Children > adults

Cluster	Size (mm^3)	*p* (peak)	*z* (peak)	x	y	z	Anatomical location[a] (probability)[b]	BA[c]
1	2,800	0.0006	3.24	68	-14	-4	Right Superior Temporal Gyrus	22
		0.0007	3.19	67.6	-16	1.2	Right Superior Temporal Gyrus	22
		0.0026	2.79	60	-14	-10	Right Middle Temporal Gyrus	21
		0.0100	2.33	54	-10	-16	Right Middle Temporal Gyrus	22
2	2,664	0.0001	3.72	2	32	40	Left Superior Medial Gyrus	8
		0.0012	3.04	-8	10	50	Left Posterior-Medial Frontal Gyrus	6
		0.0013	3.01	-6	14	50	Left Posterior-Medial Frontal Gyrus	6
		0.0035	2.70	0	20	44	Left Posterior-Medial Frontal Gyrus	6
		0.0100	2.33	0	32	50	Left Superior Medial Gyrus	8
		0.0142	2.19	0	4	58	Left Posterior-Medial Frontal Gyrus	6
3	1,928	0.0001	3.72	-52	28	0	Left Inferior Frontal Gyrus, pars triangularis (BA45-35%)	45
4	1,632	0.0014	2.99	-48	-34	8	Left Superior Temporal Gyrus	41
		0.0093	2.35	-52	-26	8	Left Superior Temporal Gyrus	41
		0.0129	2.23	-58	-22	-4	Left Middle Temporal Gyrus	21
		0.0143	2.19	-56	-40	-4	Left Middle Temporal Gyrus	21
		0.0167	2.13	-60	-22	6	Left Superior Temporal Gyrus	41
5	816	0.0057	2.53	36	16	0	Right Insula Lobe	-[d]
		0.0079	2.41	36	26	-8	Right Inferior Frontal Gyrus, pars orbitalis	13
6	680	0.0045	2.61	-48	18	22	Left Inferior Frontal Gyrus, pars triangularis (BA44-15%, BA45-7%)	9
7	488	0.0047	2.60	-32	24	-4	Left Insula Lobe	-[d]
		0.0048	2.59	-26	26	-4	Left Insula Lobe	-[d]

Adults > children

Cluster	Size (mm^3)	*p* (peak)	*z* (peak)	x	y	z	Anatomical location[a] (probability)[b]	BA[c]
1	1,000	0.0061	2.51	-48	8	8	Left Inferior Frontal Gyrus, pars opercularis (BA44-15%)	44
		0.0117	2.27	-56	10	12	Left Inferior Frontal Gyrus, pars opercularis (BA44-54%)	44

[a]according to the SPM Anatomy toolbox [b]probability of belonging to Brodmann area (BA) 44 or 45 [c]Brodmann area according to Talairach Daemon [d]no Brodmann area found by Talairach Daemon

Seed-Based Effect Size Mapping

Repeating the meta-analysis using seed-based effect size mapping, we reproduced the five clusters obtained with ALE and their respective peaks in the pars triangularis of the left IFG (BA 45), the

right insula (BA 13), bilateral MTG (BA 21), and left premotor cortex (BA 6). One additional cluster not identified by ALE emerged in the left fusiform gyrus. Furthermore, the left frontal and bilateral temporal clusters as obtained from SDM were markedly larger in size (Figure 7, Table 5).

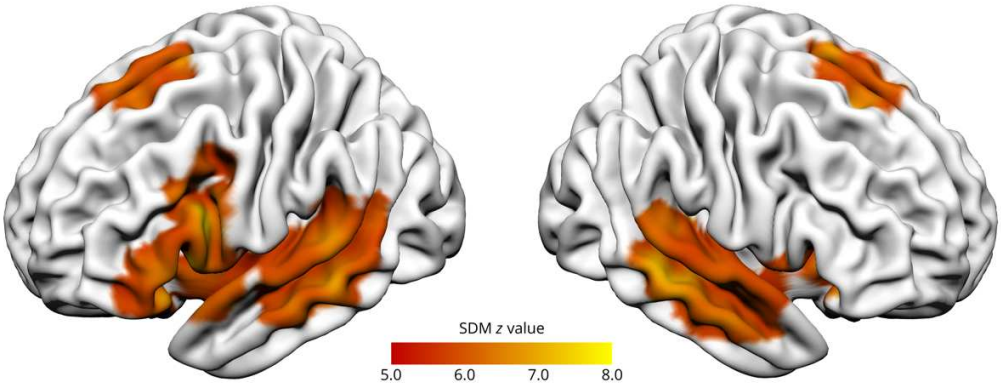


Figure 7. Significant clusters associated with language comprehension in children obtained from seed-based effect size mapping (cluster-wise FWE-corrected with a cluster-extent threshold corresponding to $p < .05$). The color bar depicts *z*-values indicating the effect size of the activation in each voxel.

Table 5

Peaks and Descriptive Information of Significant Clusters Obtained from Seed-based Effect Size Mapping

Cluster	Size (mm^3)	*p* (peak)	*z* (peak)	x	y	z	Anatomical location[a] (probability)[b]	BA[c]
1	11,392	0	8.889	-50	22	2	Left Inferior Frontal Gyrus, pars triangularis (BA45-26%, BA44-9%)	45
2	13,088	0	7.707	-66	-34	-4	Left Middle Temporal Gyrus	21
3	14,616	0	8.361	60	-28	-8	Right Middle Temporal Gyrus	21
4	5,224	0	7.196	-8	26	54	Left Superior Medial Gyrus	6
5	920	0	8.399	38	22	0	Right Insula Lobe	13
6	440	0	6.441	-32	-40	-24	Left Fusiform Gyrus	-[d]

[a]*according to the SPM Anatomy toolbox* [b]*probability of belonging to Brodmann area (BA) 44 or 45* [c]*Brodmann area according to Talairach Daemon* [d]*no Brodmann area found by Talairach Daemon*

Effects of Potentially Confounding Variables

None of the potentially confounding variables we examined (mean age of children, type of language task, type of baseline condition, and software package used for statistical analysis) were significantly related to any of the converging activation clusters for language comprehension in children. Changing the cluster-forming threshold of $p < .001$ to the extremely liberal threshold of $p < .05$, we found an effect of age in the left supplementary motor area (BA6), no effects of baseline or software package,

and an effect of task in the left supplementary motor area (BA8), left inferior temporal gyrus (BA 37), and right superior temporal gyrus (BA 21). The effect of age indicated that the older the children within an experiment, the stronger the activation in BA6. The effect of task indicated that the more complex the auditory speech stimuli used in the experiment (stories vs. sentences vs. words), the stronger the activation in BA 8, BA 37, and BA 21. However, these effects failed to reach significance at the established conservative threshold.

Jackknife Sensitivity Analysis

In the Jackknife sensitivity analysis, the five significant clusters revealed by ALE were reproduced in all 27 simulations, regardless which of the original experiments was left out (Table 6).

Table 6

Range of Results of 27 Jackknife Sensitivity Analyses

Cluster	Size (mm^3)	ALE (peak)	*z* (peak)	x	y	z	Anatomical location[a]	BA[b]
1	6,528–8,600	0.033–0.037	5.70–6.37	-50/-52	28/30	6/8	Left Inferior Frontal Gyrus, pars triangularis	45
2	4,288–5,152	0.029–0.031	5.32–5.81	-52/-56	-22/-38	2	Left Middle Temporal Gyrus	22
3	3,496–5,256	0.022–0.028	4.44–5.27	56/64	-10/-16	-8/0	Right Superior Temporal Gyrus	22
4	3,304–4,096	0.042–0.050	6.89–7.85	-2/-4	12/14	52	Left Posterior Medial Frontal Gyrus	6
5	1,352–1,936	0.020–0.026	4.09–5.03	32-38	24-30	-4/-2	Right Insula Lobe	45

[a]*according to the SPM Anatomy toolbox* [b]*Brodmann area according to Talairach Daemon*

Fail-Safe N Analysis

The fail-safe number of null experiments that could be added without altering the significance of the five clusters ranged from N = 24 for cluster 5 (right insula) to N = 115 for cluster 4 (left superior frontal gyrus; Figure 8). In each case, this number exceeded the required lower boundary of fail-safe N = 8, that is, the maximum number of null studies we estimated to be in the file drawer. Only for cluster 4 (left superior frontal gyrus), the value of fail-safe N = 115 slightly exceeded the desired upper boundary (in this case, fail-safe N = 113), potentially indicating that this cluster was driven by a very small number of experiments (Acar et al., 2018).

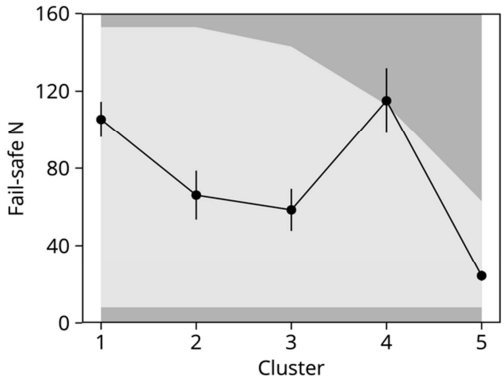


*Figure 8***.** Fail-safe N analysis for the five significant clusters associated with language comprehension in children. For every significant cluster obtained from the ALE analysis, fail-safe N indicates how many null experiments with non-significant findings could be hidden in an imaginary file drawer without compromising the statistical significance of the cluster. Light gray shading indicates the desirable fail-safe N values based on a priori considerations (see main text for details). Error bars indicate 95% confidence intervals of the mean.

Discussion

To our knowledge, here we report the first statistical synthesis of the fMRI literature on auditory language comprehension in healthy children. Meta-analyzing data reported in 24 original research articles with a total sample size of more than 600 children, we detected significant overlap in hemodynamic activation in left IFG and MTG/STG, as well as, to a lesser degree, their right-hemispheric homologues. Compared to a previous meta-analysis in adults, children revealed significantly more consistent activation in bilateral (especially right) STG and the pars triangularis and pars orbitalis of the left IFG, and significantly less consistent activation in the pars opercularis of the left IFG. In contrast to previous reviews, in which results are reported on the level of entire gyri or sulci, the present meta-analysis provides precise coordinates of consistent activation peaks in standard space. This information provides the basis for future region-of-interest studies on language processing.

According to the work of Eickhoff et al. (2016), the statistical power of the current meta-analysis to detect not only large, but also small- and medium-size effects can be assumed to be acceptable. Nevertheless, meta-analytic power is intrinsically limited by the number of currently available data (27 independent experiments). It should also be noted that most of the included individual experiments relied on sample sizes of 10 to 40 children (Figure 2B). This presumably limited their power to detect small-

and medium-size effects. These effects, in turn, were not reported as peak coordinates in the respective articles and could therefore not be included in the present analysis (Weiss-Croft & Baldeweg, 2015).

The robustness of the present results to different meta-analytic approaches was confirmed by comparing activation likelihood estimation with seed-based effect size mapping. This comparison revealed that both frameworks generated largely overlapping activation clusters. We also analyzed the robustness of the present findings to publication bias in the literature. To this end, we simulated false positives in the published literature (Jackknife sensitivity analysis) and a file drawer of unpublished studies with non-significant results (fail-safe N analysis). These analyses indicated that all of the identified clusters were robust against deleting single experiments and against adding randomly generated null experiments.

The reported differences between children and adults could be in part explained by differences in age, task, and baseline, or also to a lesser degree by the different number of studies included. While we found no evidence for significant effects of age, task, and baseline, we cannot exclude that the results are influenced by the different number of studies which is inherent to the current literature. The lack of an age effect might be explained by the age sampling variability intrinsic to the current literature. Specifically, about 60% of all studies included children with a mean age between 8 and 12 years while the age range of 3–7 years is slightly underrepresented and the age range of 13–15 years is strongly underrepresented (Figure 2A). This might have limited the statistical power of the meta-analysis to detect age-related differences.

Pooling across multiple studies, we provide evidence that the lateralization of language processing to the left hemisphere does not appear adult-like yet at a mean age of about 9 years. This finding is not in line with previous reviews stating that language lateralization is largely established by 3 to 5 years of age (Skeide & Friederici, 2016; Weiss-Croft & Baldeweg, 2015). It cannot be excluded, however, that the lack of lateralization we observed was overestimated due to the large age range of the present meta-analytic sample (3–15 years). Another explanation for this discrepancy could be that systematic reviews combine whole-brain and region-of-interest results. The current meta-analysis,

however, is entirely based on whole-brain results to ensure that different brain regions are equally likely to reveal a significant effect (Radua et al., 2012).

Our observation that children, compared to adults, recruit bilateral superior temporal cortices more consistently is in line with a large body of literature suggesting that the functional responses of the language system are not mature before young adulthood (Nuñez et al., 2011; Skeide et al., 2014; Wang et al., 2019). Specifically, children still have to rely on low-level semantic and syntactic processing implemented in the temporal cortex, while high-level semantic and syntactic processing only gradually emerges towards adulthood with an increasing involvement of the left IFG (Nuñez et al., 2011; Skeide et al., 2014; Wang et al., 2019). The notion of immature language processing in children is further corroborated by the described activation differences in the left IFG. Following our previous work, we interpret the observation that children do not yet recruit the pars opercularis to an adult-like extent as a lack of specialization of controlled syntactic processing (Nuñez et al., 2011; Skeide et al., 2014; Wang et al., 2019). Alternatively, the increasing involvement of the left pars opercularis could also be related to controlled phonological processing that refines in the course of literacy learning in school (Brennan et al., 2013). Phonological processing, however, is typically related to the dorsal pars opercularis, while in the present study, the main difference between children and adults was found in the ventral pars opercularis, a subregion that is typically related to syntactic processing (Brennan et al., 2013; Zaccarella & Friederici, 2015). Disentangling phonological, semantic, and syntactic processes during language comprehension will only be possible on a larger data basis and thus remains as a challenge for future work.

Besides the left IFG, several other regions revealed consistent activation during auditory language comprehension. Within the left temporal lobe, the left MTG is linked to the activation of lexical representations (Lau et al., 2008) and the left STG is linked to the decoding of spectro-temporal features of phonemes (Hickok & Poeppel, 2007). The right STG, in contrast, is associated with decoding supra-segmental acoustic features, i.e. the prosody of the speech input (Friederici, 2011). Within the precentral gyrus, the premotor area is thought to support language comprehension by activating subvocal articulation codes for phonemes (Pulvermüller et al., 2006). In addition to activation differences in the language

system, we found that children activated left medial and superior frontal gyri and the right insula more consistently than adults. These areas are linked to executive functions (e.g. cognitive control, performance monitoring, salience detection) and may point to the general effect that language comprehension tasks are more demanding for children than for adults (de la Vega et al., 2016; Uddin, 2015; van Noordt & Segalowitz, 2012).

Conclusion

The present meta-analysis suggests two developmental activation shifts during language comprehension that require longitudinal corroboration, namely, a triangularis-to-opercularis shift in the left inferior frontal cortex and a bilateral-to-left shift in the temporal cortex. These trajectories can be interpreted as neurodevelopmental correlates of the gradually increasing sensitivity to syntactic information.

References

Acar, F., Seurinck, R., Eickhoff, S. B., & Moerkerke, B. (2018). Assessing robustness against potential publication bias in Activation Likelihood Estimation (ALE) meta-analyses for fMRI. *PLoS ONE*, *13*(11). https://doi.org/10.1371/journal.pone.0208177

Ahmad, Z., Balsamo, L. M., Sachs, B. C., Xu, B., & Gaillard, W. D. (2003). Auditory comprehension of language in young children: Neural networks identified with fMRI. *Neurology*, *60*(10), 1598–1605. https://doi.org/10.1212/01.wnl.0000059865.32155.86

Albajes-Eizagirre, A., Solanes, A., Vieta, E., & Radua, J. (2019). Voxel-based meta-analysis via permutation of subject images (PSI): Theory and implementation for SDM. *NeuroImage*, *186*, 174–184. https://doi.org/10.1016/j.neuroimage.2018.10.077

Amunts, K., Weiss, P. H., Mohlberg, H., Pieperhoff, P., Eickhoff, S., Gurd, J. M., Marshall, J. C., Shah, N. J., Fink, G. R., & Zilles, K. (2004). Analysis of neural mechanisms underlying verbal fluency in cytoarchitectonically defined stereotaxic space—The roles of Brodmann areas 44 and 45. *NeuroImage*, *22*(1), 42–56. https://doi.org/10.1016/j.neuroimage.2003.12.031

Arredondo, M. M., Ip, K. I., Hsu, L. S. J., Tardif, T., & Kovelman, I. (2015). Brain bases of morphological processing in young children. *Human Brain Mapping*, *36*(8), 2890–2900. https://doi.org/10.1002/hbm.22815

Balsamo, L. M., Xu, B., & Gaillard, W. D. (2006). Language lateralization and the role of the fusiform gyrus in semantic processing in young children. *NeuroImage*, *31*(3), 1306–1314. https://doi.org/10.1016/j.neuroimage.2006.01.027

Bartha-Doering, L., Novak, A., Kollndorfer, K., Kasprian, G., Schuler, A.-L., Berl, M. M., Fischmeister, F. Ph. S., Gaillard, W. D., Alexopoulos, J., Prayer, D., & Seidl, R. (2018). When two are better than one: Bilateral mesial temporal lobe contributions associated with better vocabulary skills in children and adolescents. *Brain and Language*, *184*, 1–10. https://doi.org/10.1016/j.bandl.2018.06.001

Bartha-Doering, L., Novak, A., Kollndorfer, K., Schuler, A.-L., Kasprian, G., Langs, G., Schwartz, E., Fischmeister, F. Ph. S., Prayer, D., & Seidl, R. (2019). Atypical language representation is unfavorable for language abilities following childhood stroke. *European Journal of Paediatric Neurology*, *23*(1), 102–116.

https://doi.org/10.1016/j.ejpn.2018.09.007
Berl, M. M., Duke, E. S., Mayo, J., Rosenberger, L. R., Moore, E. N., VanMeter, J., Ratner, N. B., Vaidya, C. J., & Gaillard, W. D. (2010). Functional anatomy of listening and reading comprehension during development. *Brain and Language*, *114*(2), 115–125. https://doi.org/10.1016/j.bandl.2010.06.002
Berl, M. M., Mayo, J., Parks, E. N., Rosenberger, L. R., VanMeter, J., Ratner, N. B., Vaidya, C. J., & Gaillard, W. D. (2014). Regional differences in the developmental trajectory of lateralization of the language network. *Human Brain Mapping*, *35*(1), 270–284. https://doi.org/10.1002/hbm.22179
Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055
Borofsky, L. A., McNealy, K., Siddarth, P., Wu, K. N., Dapretto, M., & Caplan, R. (2010). Semantic processing and thought disorder in childhood-onset schizophrenia: Insights from fMRI. *Journal of Neurolinguistics*, *23*(3), 204–222. https://doi.org/10.1016/j.jneuroling.2009.07.004
Brauer, J., Anwander, A., & Friederici, A. D. (2011). Neuroanatomical prerequisites for language functions in the maturing brain. *Cerebral Cortex*, *21*(2), 459–466. https://doi.org/10.1093/cercor/bhq108
Brennan, C., Cao, F., Pedroarena-Leal, N., McNorgan, C., & Booth, J. R. (2013). Reading acquisition reorganizes the phonological awareness network only in alphabetic writing systems. *Human Brain Mapping*, *34*(12), 3354–3368. https://doi.org/10.1002/hbm.22147
Cao, F., Bitan, T., & Booth, J. R. (2008). Effective brain connectivity in children with reading difficulties during phonological processing. *Brain and Language*, *107*(2), 91–101. https://doi.org/10.1016/j.bandl.2007.12.009
Cao, F., Khalid, K., Lee, R., Brennan, C., Yang, Y., Li, K., Bolger, D. J., & Booth, J. R. (2011). Development of brain networks involved in spoken word processing of Mandarin Chinese. *NeuroImage*, *57*(3), 750–759. https://doi.org/10.1016/j.neuroimage.2010.09.047
Cone, N. E., Burman, D. D., Bitan, T., Bolger, D. J., & Booth, J. R. (2008). Developmental changes in brain regions involved in phonological and orthographic processing during spoken language processing. *NeuroImage*, *41*(2), 623–635. https://doi.org/10.1016/j.neuroimage.2008.02.055
de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D., & Yarkoni, T. (2016). Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. *Journal of Neuroscience*, *36*(24), 6553–6562. https://doi.org/10.1523/JNEUROSCI.4402-15.2016
Desroches, A. S., Cone, N. E., Bolger, D. J., Bitan, T., Burman, D. D., & Booth, J. R. (2010). Children with reading difficulties show differences in brain regions associated with orthographic processing during spoken language processing. *Brain Research*, *1356*, 73–84. https://doi.org/10.1016/j.brainres.2010.07.097
Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, *59*(3), 2349–2361. https://doi.org/10.1016/j.neuroimage.2011.09.017
Eickhoff, S. B., Heim, S., Zilles, K., & Amunts, K. (2006). Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *NeuroImage*, *32*(2), 570–582. https://doi.org/10.1016/j.neuroimage.2006.04.204
Eickhoff, S. B., Laird, A. R., Fox, P. M., Lancaster, J. L., & Fox, P. T. (2017). Implementation errors in the GingerALE software: Description and recommendations. *Human Brain Mapping*, *38*(1), 7–11. https://doi.org/10.1002/hbm.23342
Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based ALE meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, *30*(9), 2907–2926. https://doi.org/10.1002/hbm.20718
Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., Bzdok, D., & Eickhoff, C. R. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *NeuroImage*, *137*, 70–85. https://doi.org/10.1016/j.neuroimage.2016.04.072
Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M.-H., Evans, A. C., Zilles, K., & Amunts, K. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *NeuroImage*, *36*(3), 511–521. https://doi.org/10.1016/j.neuroimage.2007.03.060
Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034
Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, *29*(5), 581–593. https://doi.org/10.1002/hbm.20422
Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*,

91(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011
Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. JSTOR. https://doi.org/10.2307/1164588
Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113
Holland, S. K., Vannest, J., Mecoli, M., Jacola, L. M., Tillema, J.-M., Karunanayaka, P. R., Schmithorst, V. J., Yuan, W., Plante, E., & Byars, A. W. (2007). Functional MRI of language lateralization during development in children. *International Journal of Audiology*, *46*(9), 533–551. https://doi.org/10.1080/14992020701448994
Horowitz-Kraus, T., Buck, C., & Dorrmann, D. (2016). Altered neural circuits accompany lower performance during narrative comprehension in children with reading difficulties: An fMRI study. *Annals of Dyslexia*, *66*(3), 301–318. https://doi.org/10.1007/s11881-016-0124-4
Horowitz-Kraus, T., DiFrancesco, M., Kay, B., Wang, Y., & Holland, S. K. (2015). Increased resting-state functional connectivity of visual- and cognitive-control brain networks after training in children with reading difficulties. *NeuroImage. Clinical*, *8*, 619–630. https://doi.org/10.1016/j.nicl.2015.06.010
Hubbard, A. L., McNealy, K., Scott-Van Zeeland, A. A., Callan, D. E., Bookheimer, S. Y., & Dapretto, M. (2012). Altered integration of speech and gesture in children with autism spectrum disorders. *Brain and Behavior*, *2*(5), 606–619. https://doi.org/10.1002/brb3.81
Knoll, L. J., Obleser, J., Schipke, C. S., Friederici, A. D., & Brauer, J. (2012). Left prefrontal cortex activation during sentence comprehension covaries with grammatical knowledge in children. *NeuroImage*, *62*(1), 207–216. https://doi.org/10.1016/j.neuroimage.2012.05.014
Lancaster, J. L., Rainey, L. H., Summerlin, J. L., Freitas, C. S., Fox, P. T., Evans, A. C., Toga, A. W., & Mazziotta, J. C. (1997). Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method. *Human Brain Mapping*, *5*(4), 238–242. https://doi.org/10.1002/(SICI)1097-0193(1997)5:4<238::AID-HBM6>3.0.CO;2-4
Lancaster, Jack L., Tordesillas□Gutiérrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J. C., & Fox, P. T. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping*, *28*(11), 1194–1205. https://doi.org/10.1002/hbm.20345
Lancaster, Jack L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., Kochunov, P. V., Nickerson, D., Mikiten, S. A., & Fox, P. T. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping*, *10*(3), 120–131. https://doi.org/10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8
Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532
Monzalvo, K., & Dehaene-Lambertz, G. (2013). How reading acquisition changes children's spoken language network. *Brain and Language*, *127*(3), 356–365. https://doi.org/10.1016/j.bandl.2013.10.009
Monzalvo, K., Fluss, J., Billard, C., Dehaene, S., & Dehaene-Lambertz, G. (2012). Cortical networks for vision and language in dyslexic and normal children of variable socio-economic status. *NeuroImage*, *61*(1), 258–274. https://doi.org/10.1016/j.neuroimage.2012.02.035
Moore-Parks, E. N., Burns, E. L., Bazzill, R., Levy, S., Posada, V., & Müller, R.-A. (2010). An fMRI study of sentence-embedded lexical-semantic decision in children and adults. *Brain and Language*, *114*(2), 90–100. https://doi.org/10.1016/j.bandl.2010.03.009
Nuñez, S. C., Dapretto, M., Katzir, T., Starr, A., Bramen, J., Kan, E., Bookheimer, S., & Sowell, E. R. (2011). fMRI of syntactic processing in typically developing children: Structural correlates in the inferior frontal gyrus. *Developmental Cognitive Neuroscience*, *1*(3), 313–323. https://doi.org/10.1016/j.dcn.2011.02.004
Pulvermüller, F., Huss, M., Kherif, F., Martin, F. M. del P., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(20), 7865–7870. https://doi.org/10.1073/pnas.0509989103
Radua, J., Mataix-Cols, D., Phillips, M. L., El-Hage, W., Kronhaus, D. M., Cardoner, N., & Surguladze, S. (2012). A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *European Psychiatry*, *27*(8), 605–611. https://doi.org/10.1016/j.eurpsy.2011.04.001
Radua, Joaquim, & Mataix-Cols, D. (2012). Meta-analytic methods for neuroimaging data explained. *Biology of Mood & Anxiety Disorders*, *2*, 6. https://doi.org/10.1186/2045-5380-2-6
Raschle, N. M., Smith, S. A., Zuk, J., Dauvermann, M. R., Figuccio, M. J., & Gaab, N. (2014). Investigating the Neural Correlates of Voice versus Speech-Sound Directed Information in Pre-School Children. *PLoS ONE*, *9*(12). https://doi.org/10.1371/journal.pone.0115549
Redcay, E., Haist, F., & Courchesne, E. (2008). Functional neuroimaging of speech perception during a pivotal

period in language acquisition. *Developmental Science*, *11*(2), 237–252. https://doi.org/10.1111/j.1467-7687.2008.00674.x

Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An Activation Likelihood Estimation meta-analysis. *Brain and Language*, *141*, 89–102. https://doi.org/10.1016/j.bandl.2014.11.012

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. E. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science*, *29*(5), 700–710. https://doi.org/10.1177/0956797617742725

Samartsidis, P., Montagna, S., Laird, A. R., Fox, P. T., Johnson, T. D., & Nichols, T. E. (2019). Estimating the prevalence of missing experiments in a neuroimaging meta-analysis. *BioRxiv*, 225425. https://doi.org/10.1101/225425

Skeide, M. A., Brauer, J., & Friederici, A. D. (2014). Syntax gradually segregates from semantics in the developing brain. *NeuroImage*, *100*, 106–111. https://doi.org/10.1016/j.neuroimage.2014.05.080

Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, *17*(5), 323–332. https://doi.org/10.1038/nrn.2016.23

Sroka, M. C., Vannest, J., Maloney, T. C., Horowitz-Kraus, T., Byars, A. W., Holland, S. K., & CMIND Authorship Consortium. (2015). Relationship between receptive vocabulary and the neural substrates for story processing in preschoolers. *Brain Imaging and Behavior*, *9*(1), 43–55. https://doi.org/10.1007/s11682-014-9342-8

Szaflarski, J. P., Holland, S. K., Schmithorst, V. J., & Byars, A. W. (2006). An fMRI study of language lateralization in children and adults. *Human Brain Mapping*, *27*(3), 202–212. https://doi.org/10.1002/hbm.20177

Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage*, *16*(3 Pt 1), 765–780. https://doi.org/10.1006/nimg.2002.1131

Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012). Minimizing within-experiment and within-group effects in Activation Likelihood Estimation meta-analyses. *Human Brain Mapping*, *33*(1), 1–13. https://doi.org/10.1002/hbm.21186

Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*, *16*(1), 55–61. https://doi.org/10.1038/nrn3857

van Noordt, S. J. R., & Segalowitz, S. J. (2012). Performance monitoring and the medial prefrontal cortex: A review of individual differences and context effects as a window on self-regulation. *Frontiers in Human Neuroscience*, *6*, 197. https://doi.org/10.3389/fnhum.2012.00197

Vannest, J., Maloney, T. C., Tenney, J. R., Szaflarski, J. P., Morita, D., Byars, A. W., Altaye, M., Holland, S. K., & Glauser, T. A. (2019). Changes in functional organization and functional connectivity during story listening in children with benign childhood epilepsy with centro-temporal spikes. *Brain and Language*, *193*, 10–17. https://doi.org/10.1016/j.bandl.2017.01.009

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30*(4), 1414–1432. https://doi.org/10.1016/j.neuroimage.2005.11.002

Wang, Z., Yan, X., Liu, Y., Spray, G. J., Deng, Y., & Cao, F. (2019). Structural and functional abnormality of the putamen in children with developmental dyslexia. *Neuropsychologia*, *130*, 26–37. https://doi.org/10.1016/j.neuropsychologia.2018.07.014

Weiss-Croft, L. J., & Baldeweg, T. (2015). Maturation of language networks in children: A systematic review of 22years of functional MRI. *NeuroImage*, *123*, 269–281. https://doi.org/10.1016/j.neuroimage.2015.07.046

Zaccarella, E., & Friederici, A. D. (2015). Merge in the human brain: A sub-region based functional investigation in the left pars opercularis. *Frontiers in Psychology*, *6*, 1818. https://doi.org/10.3389/fpsyg.2015.01818