Published in final edited form as:
Neuroimage. Vol 62, Iss 3, Pages 1987–98, 2012 DOI: 10.1016/j.neuroimage.2012.05.052

Linking ordering in Broca's area to storage in left temporo-parietal regions: The case of sentence processing

Lars Meyer [a,*], Jonas Obleser [b], Alfred Anwander [a], Angela D. Friederici [a]

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

[b] Max Planck Research Group "Auditory Cognition", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Corresponding author:
Lars Meyer
Max Planck Institute for Human Cognitive and Brain Sciences,
Department of Neuropsychology,
Stephanstrasse 1a
04103 Leipzig
Germany
http://www.cbs.mpg.de/~lmeyer

*Preprint submitted to **Neuroimage** (2012).* ***Final draft: May 2012***

Abstract

In sentence processing, storage and ordering of the verb and its arguments (subject and object) are core tasks. Their cortical representation is a matter of ongoing debate, and it is unclear whether prefrontal activations in neuroimaging studies on sentence processing reflect the storage of arguments or their ordering. Moreover, it is unclear how storage during sentence processing relates to the neuroanatomy of storage outside the sentence processing domain. To tackle these questions, we crossed the factor "ordering" (subject-first vs. object-first German sentences) with the factor "storage" (one vs. four phrases intervene between the critical argument and the verb) in an auditory fMRI study. Ordering focally activated the left pars opercularis in Broca's area, while storage activated deep left temporo-parietal (TP) regions. Notably, left TP activation correlated with listener's digit span, while Broca's area activation did not. Furthermore, fractional anisotropy of listeners' left arcuate fasciculus/superior longitudinal fasciculus (AF/SLF) is shown to covary with the functional effect of increased storage demands at sites along the tract. Functionally, the results suggest that storage during sentence processing relies on TP regions, likely shared between sentence processing and other working memory-related tasks, while Broca's area appears as a distinct neural correlate of ordering. We conclude that the abstract notion of sentence processing can be captured by the interplay of concrete cognitive concepts such as ordering and storage.

Keywords: inferior frontal gyrus, inferior parietal cortex, sentence processing, superior longitudinal fasciculus, working memory

1 Introduction

The core task in sentence processing is to determine who is doing what to whom. To succeed, the who and whom—the arguments, i.e., subject and object—must be linked to the doing—the verb. The minimal neurocognition of this task has to account for two aspects of processing: storage and ordering processes. First, the arguments have to be stored in working memory until they can be retrieved for integration with the verb, in particular when the verb is in the sentence-final position. Second, in case of word order manipulations the arguments have to be ordered, such that the who and the whom are not confused. There is firm evidence for the neuropsychological reality of both storage and ordering during argument–verb dependency processing from behavioral and event-related brain potential (ERP) studies (Clahsen and Featherston, 1999; Felser et al., 2003; Fiebach et al., 2001; Kluender and Kutas, 1993; Nakano et al., 2002; Nicol et al., 1994; Nicol and Swinney, 1989; Phillips et al., 2005; Ueno and Kluender, 2003). Nevertheless, the neural substrate of storage and ordering is one of the most enduring issues in the neuroscience of language.

With respect to storage, an exemplary imaging study from sentence processing reports activity of left inferior parietal cortex to increase with the retention interval for disambiguating information in ambiguous sentences, that is, storage demands (Novais-Santos et al., 2007). In addition, meta-analyses and imaging studies from outside of sentence processing suggest left posterior brain regions as candidates for a storage substrate (Owen et al., 2005; Smith and Jonides, 1998; Wager et al., 2005). Paulesu et al's (1993) seminal study required monolingual English-speaking participants to read, store and rehearse either English or Korean letters, of which only the former were hypothesized to activate a phonological code and subsequent storage. This study found the left supramarginal gyrus (SMG) to be active during the storage of English letters, a finding which was replicated by later studies (Awh et al., 1996; Awh et al., 1995; D'Esposito et al., 1999; Jonides et al., 1998; Petrides et al., 1993). However, these studies contrast with other imaging work investigating aspeects of working memory during sentence processing that reports Brodmann Area (BA) 45, sometimes extending to the inferior frontal sulcus, in the left prefrontal cortex to play a role, using paradigms comparing different syntactic dependencies (Fiebach et al., 2005; Makuuchi, et al. 2009; Santi and Grodzinsky 2007; Santi and Grodzinsky 2010).

With respect to ordering during sentence processing, Broca’s area in the inferior frontal cortex has been found to activate in languages as diverse as Hebrew, German and Japanese (Ben-Shachar et al., 2003; Bornkessel et al., 2005; Friederici et al., 2006b; Kim et al., 2009). Previous studies investigating working memory did not explicitely separate ordering and storage. One study directly contrasted word order and distance between the verb and its arguments during sentence processing, showing the certain independence of these two factors within the inferior frontal cortex, but restricted to specific types of syntactic dependencies (Fiebach et al., 2005). Another study neuroanatomically separated processing of syntactic hierarchies in Broca’s area from working memory in the left inferior frontal sulcus, but this study did not vary ordering demands (Makuuchi et al., 2009).

Some authors have claimed that the role of Broca’s area in sentence processing is not in ordering, but rather in subvocal rehearsal of stored information (Rogalsky et al., 2008; Rogalsky and Hickok, 2010). Although Broca’s area was active during subvocal rehearsal in Paulesu et al.’s (1993) and subsequent non-sentence processing studies (Awh et al., 1996; Awh et al., 1995; Petrides et al., 1993), no previous imaging data could disentangle ordering and rehearsal during sentence processing. Although Rogalsky et al. (2008) report a decline in sentence processing performance under conditions of articulatory rehearsal, they found a control condition (finger tapping during a sentence processing task) to also selectively decrease sentence processing performance, merely suggesting that the presence of a secondary task can affect sentence processing behaviorally. The claim that activation in Broca's area during sentence processing stems from subvocal rehearsal is further questioned by clinical evidence that ordering in sentence processing can be independently impaired from active subvocal rehearsal (Caplan and Waters, 1999; Waters and Caplan, 1996). In addition, there is imaging evidence that subvocal rehearsal during sentence processing does not further increase brain activation in Broca's area (Caplan et al., 2000). In sum, while both rehearsal and ordering may rely on Broca's area, a conceptual and neuroanatomical identity between ordering and rehearsal is not straightforward.

Based on the above findings, we hypothesize that the minimal neural representation of subject, object and verb involves at least two computational components: a storage component which stores arguments across the argument–verb distance, and an ordering component which orders these arguments to avoid a who–whom confusion. We predict that ordering is supported by Broca's area, and that the storage of relevant features of an argument over the argument–verb distance is supported by temporo-parietal regions. This hypothesis is tested in the present study, which fully crossed the factors "ordering" (i.e., argument order) and "storage" (i.e., argument–verb distance) in a functional magnetic resonance imaging (fMRI) investigation. Importantly, if storage rather than subvocal rehearsal supports sentence processing, we hypothesize to find temporo-parietal (TP) rather than prefrontal brain activation for increased argument–verb distance. Moreover, if TP regions subserve storage and Broca's area subserves ordering, these regions should exhibit a direct fiber connection, and the local microstructural properties of the fiber bundle should be related to the fMRI activations. To evaluate this, we also conducted an analysis of diffusion MRI data using deterministic tractography followed by a voxel-based statistic on the characteristic diffusion parameters.

2 Methods

2.1 Participants

Twenty-four university students (mean age 27.1 years, standard deviation (SD) 3.2 years, 12 females, all native speakers of German) took part in the experiment. Participants were matched for their reading span being in the range between 3 and 5 (mean 3.89, SD 0.77) according to an abridged version of the reading span test (Daneman and Carpenter, 1980). All of them were right-handed as assessed by an abridged version of the Edinburgh Inventory (Oldfield, 1971), had no reported neurological or hearing deficits, and normal or corrected to normal vision. Participants were paid €14 for participating. Written informed consent was obtained from all subjects. All procedures received ethical approval by the local ethics committee (University of Leipzig).

2.2 Working memory test

Participants' verbal working memory abilities for sequences of items was tested in two tests on digit span by using a German version of the respective subtest from the Wechsler test (Tewes, 1994); mean forward digit span was 9.54 (SD = 1.69), mean backward digit span was 8.33 (SD = 2.10).

2.3 Sentence Materials

German sentences were constructed to allow for orthogonal manipulation of one factor solely affecting ordering and one factor solely affecting storage. The 2 × 2 factorial design accordingly crossed the factors ordering (subject-first vs. object-first) and storage (short vs. long), as shown in Figure 1.

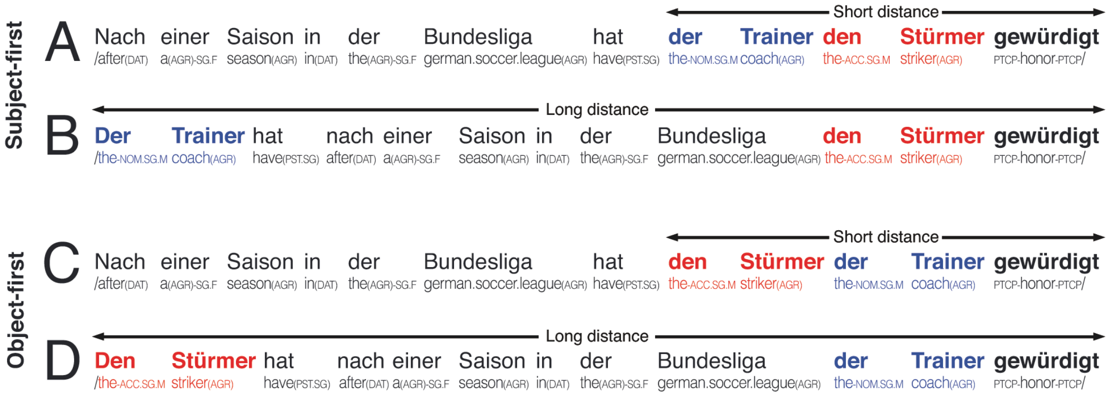

Figure 1: Overview of experimental materials; upper panel shows subject-first orders in the short (A) and long (B) distance variants, lower panel shows object-first orders in the short (C) and long (D) distance variants; interlinear glosses are provided below each example. The subject of each sentence is marked in bold blue, each sentence's object is marked in bold red, the main verb ist marked in bold black font. Arrows illustrate the argument–verb distance, operationalized as the distance between the critical argument and the main verb. All sentences translate *After a season in the german soccer league, the coach honored the striker.*

In the first condition (A), subject and object are in the immediate vicinity of the main verb—the argument order is subject-first, and the argument–verb distance is short. In the second condition (B), the argument order is still subject-first, while the subject is at the sentence beginning now. This increases storage demands by lengthening the retention interval for the critical information of the subject noun phrase. The third (C) and fourth (D) conditions involve the same two distance variants, but in a different argument order: Here, the object noun phrase is in front of the subject noun phrase, again lenghtening the retention interval. These stimuli have the crucial advantage that storage demands are identical for the two short and the two long conditions, respectively (that is, in the short conditions, one phrase intervenes between the first argument and the main verb, whereas in the long conditions, four phrases intervene). In addition, the absolute position of the main verb is kept constant across conditions to avoid the effect that processing speed increases towards sentence ending (Ferreira and Henderson, 1993). In order to avoid the sentence-final processing slowdown due to increasing propositional and semantic load (Friedman et al., 1975), we added a conjunct clause (e.g., ”und die Entwicklung bestätigt.”, translating to *and stated the development.* for the examples in Figure 1). This conjunct clause was identical across all four experimental conditions of a set. Furthermore, to avoid any influence of frequency and semantic coherence on the dependent variables (Van Petten, 1993), a position-wise lemma frequency and syllable count matching using the CELEX database (Baayen et al., 1995) was performed and supplemented by sentential neighborhood analyses using the Projekt Deutscher Wortschatz database (Biemann et al., 2004). Specifically, subject and object of each sentence were balanced in length and frequency to avoid systematic confounding of the ordering manipulation. According to these criteria, 48 sentences in the four conditions were made, resulting in a set of 192 stimuli. In addition, 192 sentences (e.g., ”Gestern hat der Leser dem Bibliothekar den Artikel zurückgegeben”, translating *Yesterday, the reader gave back the article to the librarian*) of differing ordering and storage demands from a previous study (Friederici et al., 2006b) served as filler items to avoid multicolinearity in the design (Andrade et al., 1999). In order to fit these sentences to the experimental stimuli, conjunct clauses (e.g., ”und die Gebühr bezahlt.”, translating *and paid the fees.*) were added as well, position-wise frequency-matched to the experimental stimuli using the CELEX database (Baayen et al., 1995). This resulted in a final stimulus pool of 384 sentences.

All sentences were recorded in a soundproof chamber by a trained female German speaker with a Sennheiser® MKH 40 condenser microphone and a Roland® CD-2 digital sound recorder. The recordings were cut and normalized in Praat (Boersma and Weenink, 2001) according to the root mean square amplitude of all files. To avoid onset and offset artifacts, a cosine fade in and out sequence of 5 ms was attached. For each subject, an individual pseudo-randomized list of 216 stimuli was generated using MATLAB® (The MathWorks, Inc., Natick, MA, USA) scripts. A list consisted of 144 stimulus sentences (36 per condition), 36 filler sentences and 36 null events. These were drawn in a counterbalanced way from the entire stimulus pool. As a task to maintain participants' attention, twenty-four of the stimulus sentences additionally introduced a who-did-what-to-whom yes/no comprehension question to be answered within a limited time (e.g., *Hat der Trainer den Stürmer geehrt? / Did the coach honor the center forward?*). Comprehension questions appeared in 16.7 % of trials in each of the four conditions at unpredictable positions in the stimulus set. The proportion of yes–correct and no–incorrect questions was balanced.

2.3 Procedure

Stimuli were presented using the software package Presentation® (Neurobehavioural Systems, Inc., Albany, CA, USA). Auditory stimuli were presented using airconduction headphones (Resonance Technology, Inc., Northridge, CA, USA). Visual stimuli were presented on a Sanyo PLC-XP50L LCD XGA mirror-projection system with a refresh rate of 100 Hz (Sanyo Electric Co., Ltd., Moriguchi, Japan), mounted onto the headcoil. A sans-serif font in black letters against a gray background (font size 20 px) was used.

A trial started with a fixation cross that stayed on screen for the whole trial. After a random jitter of either 0, 500, 1000 or 1500 ms, an auditory stimulus started (mean length 4.9 s, SD 0.36 s). To keep the number of acquired volumes constant across trials, a trial always lasted for 8 s, interpolating a silent period and an on-screen fixation cross between stimulus and trial end. In the case of a comprehension question, such a sequence was followed by a fixation cross of a random jitter and a visual comprehension question (16.7 % of all trials). The question remained on screen for 1500 ms and had to be answered by the participant as quickly as possible during this time period. Subsequent visual feedback was given for 1000 ms by a green happy or red sad emoticon. Again, in order to keep the duration of the comprehension probes constant, silence and an on-screen fixation cross were interpolated, such that each comprehension probe would last 4 s. Participants were instructed to carefully listen to the sentences and to answer the comprehension questions via button press with either their left or right hand, with one hand corresponding to yes and the other to no. Response button assignment was counterbalanced across participants. Since participants were not aware of whether a task trial would occur during the auditory stimulus, all trials were included in the analysis.

2.4 MR acquisition parameters

Functional image acquisition was performed with a 3 Tesla Siemens TIM TRIO scanner (Siemens Healthcare, Erlangen, Germany) and a 12-channel headcoil at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. T1-weighted 3D MP-RAGE (magnetization-prepared rapid gradient echo) images (Mugler III and Brookeman, 1990; TI = 650 ms; TR = 1300 ms; alpha = 10°; FOV = 256 mm x 240 mm; 2 acquisitions; 1 mm isotropic resolution) were previously acquired with a non-slice-selective inversion pulse followed by a single excitation of each slice, and were available for preprocessing of the functional data. As to the functional data acquisition, a T2*-weighted gradient-echo echo-planar-imaging (EPI) sequence was run (data matrix 64 × 64, TR = 2.0 s, continuous scanning, TE = 30 ms, flip angle = 90°, bandwidth 116 kHz, FOV = 19.2 cm, in-plane resolution 3 mm × 3 mm, slice thickness 3 mm, interslice gap 1 mm, 30 horizontal slices parallel to AC-PC line, whole-brain coverage, 912 volumes), with a functional scan time of 30 min.

Diffusion-weighted data were acquired in a separate session with a 3 Tesla Siemens TIM TRIO scanner (Siemens Healthcare, Erlangen, Germany), equipped with a 32-channel phased-array head array coil. Images were acquired with a twice-refocused spin echo EPI sequence (Reese et al., 2003; TE = 100 ms, TR = 12 s, 128 × 128 image matrix, FOV = 220 × 220 mm^2, 88 axial slices (no gap), resolution: 1.72 × 1.72 × 1.7 mm^3). Additionally, fat saturation was employed together with 6/8 partial Fourier imaging and generalized auto-calibrating partially parallel acquisitions (GRAPPA, acceleration factor = 2; Griswold et al., 2003). Diffusion-weighting was isotropically distributed along 60 diffusion-encoding gradient directions with a b-value of 1000 s/mm^2. Seven images with no diffusion-weighting (b0) were acquired initially and interleaved after each block of 10 diffusion-weighted images as anatomical reference for offline motion correction. The diffusion MRI sequence lasted about 16 minutes.

2.5 Data Analysis

For the behavioral data, we calculated d'-scores and reaction times. Although mean percentage correct scores are often used, d'-scores are a more adequate representation of participants' performance in that they efficiently eliminate participants' response bias, that is, the individual tendency to either press the yes–correct or no–incorrect button (Macmillan and Creelman, 2005). A one-sample t-test on the difference between the d'-scores and chance level performance (50 % correct responses) was performed. A 2 × 2 Analysis of Variance (ANOVA) was run on the response data to determine condition-specific effects. From the individual scores on the forward and backward digit span subtests, we calculated a composite score—the Mahalanobis distance—based on a zero-centered sample with covariance corresponding to our sample (Tabachnik et al., 2001).

2.5.1 Functional MRI preprocessing and whole-head statistics

Functional data analysis was performed using the SPM 8 software package (Wellcome Imaging Department, University College, London, UK). Before undergoing statistical analysis, the functional data were co-registered using the corresponding high-resolution 3-D structural images. They were resampled to 3 × 3 × 3 mm^3 voxel size. Further preprocessing was performed by realigning the functional time series to the first image, correcting them for slice timing and field inhomogeneities. Next, normalization to the standard MR template (gray-matter segmentation-based procedure) and smoothing using an isotropic 8 mm^3 kernel were applied. For statistical analysis, a participant-wise general linear model was estimated using the canonical hemodynamic response function from the SPM 8 software package, starting at sentence onset and spanning the individual stimulus length (mean 4.89 s, SD 0.36 s), and treating fillers, silent trials and comprehension questions as regressors of no interest. In addition, we included the performance measures for each condition (see Results) as a regressor of no interest in order to factor out performance-related variance from our analysis. A high pass filter of 1/100 s was used to attenuate slow global signal changes. Contrast estimates for the four experimental conditions (compared against the global mean) were passed into a second-level within-subject ANOVA, where main effects and interactions were assessed. For thresholding of the statistical parametrical maps, an AFNI-implemented Monte Carlo simulation (NIMH Scientific and Statistical Computing Core, Bethesda, MD, USA) ensured that a cluster extent of at least 51 voxels and an uncorrected voxel-wise p-value of 0.005 would protect against whole-volume type I error at $\alpha = 0.05$. For anatomical assessment of functional activations, cytoarchitectonic maps provided with the SPM anatomy toolbox (Eickhoff et al., 2005) were used.

2.5.2 Correlation Analysis

Using the Marsbar toolbox (Brett et al., 2002), the linear predictor values for each participant and design cell for the resulting group-peak activation-clusters (regions of interest; ROIs) were transformed into percentages of signal change. From these, difference values for the two main manipulation effects (i.e., the ordering and storage effect) were computed and Pearson's linear correlations between the individual activations and Mahalanobis-transformed composite digit span scores were run.

2.5.3 Diffusion-weighted MRI data analysis

For all 22 participants who remained in the analysis (see Results), diffusion MRI data were available and were analyzed using LIPSIA (Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany), FSL (FMRIB, University of Oxford, United Kingdom) and SPM 8 (Wellcome Imaging Department, University College, London, UK) on a Linux workstation. T1-weighted structural scans were used for skull stripping and the brain images were co-registered into Talairach space (Talairach and Tournoux, 1988). Motion correction for the 60 diffusion-weighted images was performed based on the 7 reference images without diffusion weighting distributed over the entire sequence using rigid-body transformations (Jenkinson et al., 2002) implemented in FSL. Motion correction parameters were interpolated for all 67 volumes and combined with a global registration to the T1 anatomy, which was computed with the same method. The transformations were applied for all volumes resulting in an isotropic voxel resolution of 1 mm. The gradient direction for each volume was corrected using the rotation parameters. Finally, for each voxel, a diffusion tensor model (Basser et al., 1994) was fitted to the raw data and the fractional anisotropy (FA) was computed (Basser and Pierpaoli, 1996). The FA parameterizes the local diffusion profile and is sensitive to differences in local connectivity of individual subjects (Brauer et al., 2011).

To robustly analyze a cross-subject correlation between the FA and the signal change in the functional activation, we analyzed the FA values within the skeleton of the fiber bundle connecting the activated areas. Following the approach of tract-based spatial statistics (TBSS; Smith et al., 2006) implemented in FSL, we normalized all FA images into a standard brain space and computed a group-average FA image. A mean FA skeleton was created from this image, which represents the centers of all tracts common to the group. The local maxima of the individual aligned FA data representing the individual centers of the tracts were then projected onto this skeleton.

Anatomical connectivity between the functionally activated areas was investigated by tractography from the diffusion tensor maps to estimate the location of the corresponding fiber bundle within the individual brain volume (Anwander et al., 2007). Here we used deterministic tractography using the entire diffusion tensor to deflect the estimated fiber trajectory (Lazar et al., 2003) as implemented in MedINRIA (Asclepios, INRIA, Sophia Antipolis, France) according to Fillard et al. (2007). We used the same preprocessing chain as for the computation of the FA maps mentioned above, except for the fact that the diffusion tensors were computed with an isotropic resolution of 1.7 mm. Fiber trajectories were started in all voxels with an FA > 0.2 resulting in a complete set of trajectories within the brain.

To use the functional activation clusters (see Results) as symmetric seed and target regions in deterministic fiber tracking, the group average statistical maps were back-projected into each participant's native image space, and all fiber trajectories connecting the two regions in each individual participant were selected as white matter connections. The resulting individual tract volume was mapped onto the voxel space of the individual anatomical scan. These tractograms were projected onto the group-average FA data set using the transformation matrices obtained in the TBSS analysis. Voxels inside these normalized tracts were included in an individual logical map, and these maps were summed across participants to obtain a group-level probability map. The probability map was thresholded at $p < 0.5$ (i.e. at least 50 % of participants had a tract at this voxel) to leave only the core volume intact (Reich et al., 2010). This provided an across-participants search volume connecting the functional activation clusters and only representing white-matter voxels present across participants.

This volume was used to mask a voxelwise multiple regression analysis in SPM on the aligned individual data as resulting from the TBSS procedure. Individual percent signal change from the two functional clusters (see Results) was used as regressors, correcting for total intracranial volume, age and gender. The volume for the statistical analysis encompassed only voxels inside the final tract volume (1561 voxels), and an AFNI-implemented Monte Carlo simulation (NIMH Scientific and Statistical Computing Core, Bethesda, MD, USA) ensured that a cluster of at least 7 voxels at a voxel-wise p-value of 0.05 protected against type I errors at $\alpha = 0.005$. All stereotaxic coordinates are reported in standard Montreal Neurological Institute (MNI) space.

3 Results

3.1 Behavioral data

Mean d'-score across conditions was 0.66 (SD = 0.94) with a mean response bias of c = –0.22 (SD = 0.59), and mean reaction time was 997 ms (SD = 155 ms). One participant was excluded from all further analyses because his average response time of 450 ms was outside of the 95–% confidence interval for the group. Although the d'-scores indicate serious task challenges for the listeners, a one-sample t-test on these scores showed that participants' performance was significantly better than chance (i.e., a d' of zero; $t(23) = 3.49$, $p < 0.005$). Furthermore, the 2 × 2 ANOVA on neither the condition-specific scores (Figure 2A) nor reaction times (Figure 2B) yielded any main effects or interactions ($p > 0.2$). This indicates that our design was free of processing-difficulty confounds.

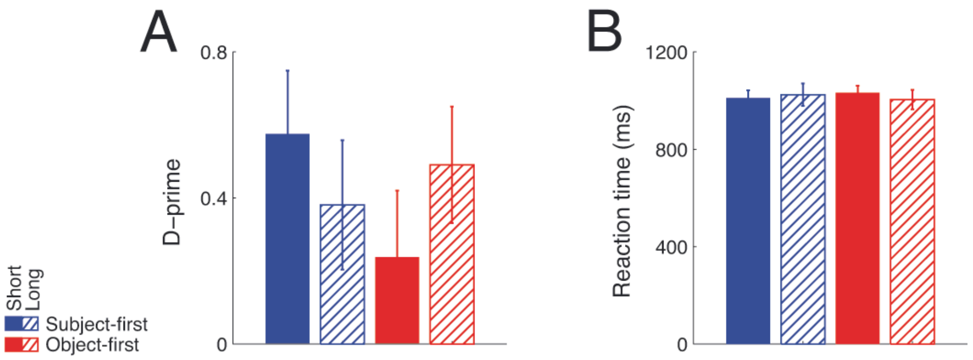

Figure 2: (A) D'-scores and (B) reaction times for all four conditions; error bars mark standard error of the mean (SEM).

3.2 Functional MRI results

For both main effects in our design, that is, ordering and storage, focal and exclusive supra-threshold activations in the left hemisphere were obtained. As shown in the panels A and B of Figure 3 (shown in red), a test for a main effect of ordering elicited activation in the left pars opercularis of the left inferior frontal gyrus (IFG), peaking at x = –54, y = 10, z = 18. According to the cytoarchitectonic probability maps provided by Eickhoff et al. (2005), 64.8 % of the obtained activation mapped onto Brodmann area (BA) 44, while a smaller part of 19.5 % mapped onto BA 45. Panels B and C of Figure 3 (shown in blue) illustrate the activation for the main effect of storage, which peaked at x = –42, y = –40, z = 10. This activation peak maps onto lateral superior parts of the left Planum Temporale (PT) with an across-participant probability of 26–45 %, according to probability maps provided by Westbury et al. (1999). Since the activation cluster encompassed regions both in the temporal and inferior parietal region, we refer to it as TP activation. The interaction between the ordering and storage factor did not produce activations that survived the statistical thresholds; specifically, there was no statistical evidence for prefrontal activation by the storage manipulation. The full set of activations is shown in Table 1.

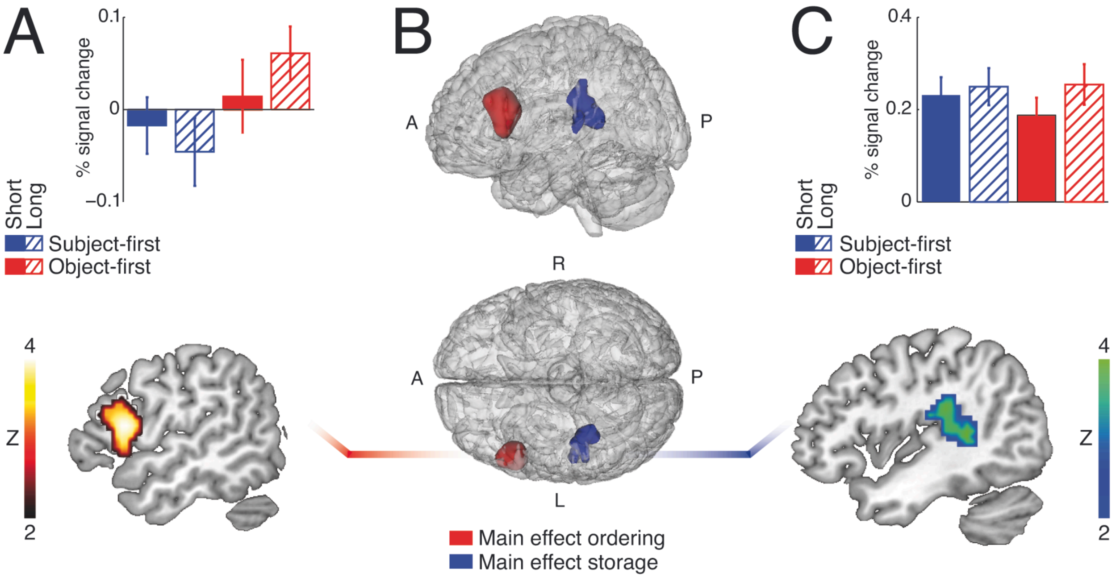

Figure 3: Brain activations and signal change (bar plots including SEM) for (A) the ordering effect (red clusters in panel B) and (C) the storage effect (blue clusters in panel B). All activations are thresholded at $p < 0.005$ at a minimum cluster size of 51 suprathreshold voxels to achieve type I error control at $p < 0.05$. For the ordering factor, 64.8 % of the obtained cluster is in the left BA 44, more specifically the left IFG, with a peak at $x = –54$, $y = 10$, $z = 18$ ($z = 3.76$). For the storage factor, the activation is found in the medial left PT, peaking at $x = –42$, $y = –40$, $z = 10$ ($z = 3.66$).

Table 1: Overview of significant clusters in the functional contrasts surviving the 51-voxel threshold at $p < 0.005$ to achieve whole-volume type I error control at $p < 0.05$.

Site	MNI-coordinate			Cluster size	Z-score
	X	Y	Z	(mm^3)	
Object-first > subject-first					
Left BA 44 / pars opercularis*	–54	14	13	1602	3.76
Long > short					
Left planum temporale**	–42	–40	10		3.66
Left supramarginal gyrus***	–45	–28	22	1485	3.27
Left superior temporal gyrus***	–57	–25	1		2.89

*According to Eickhoff et al. (2005), **according to Westbury et al. (1999), ***according to Talairach and Tournoux (1988).

Notably, a Pearson’s linear correlation analysis found that the difference in signal change in the left TP ROI (as defined by the functional effect in the TP region) was negatively correlated with the combined digit span scores ($r = -0.45$, $p < 0.05$). Participants with higher verbal short-term memory scores show relatively less signal change in the contrast reflecting the storage effect (Figure 4A). No other correlations were significant. In particular, there was no substantial correlation between activity in the IFG in the storage contrast and the combined digit span scores ($r = -0.22$, $p > 0.3$; see Figure 4B). In sum, individual combined digit spans were able to explain $r^2 \sim 20$ % of the storage effect variance in the TP region, but only $r^2 \sim 4$ % of the analogue effect in the IFG.

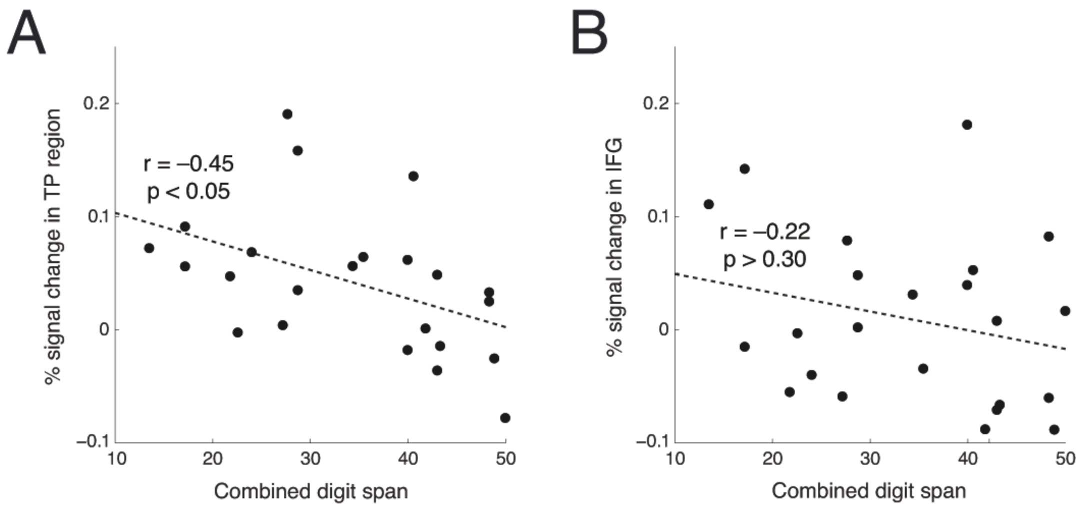

Figure 4: (A) Negative correlation of storage effect in the medial left PT with combined digit span (left, r = –0.45, p < 0.05); (B) analogue correlation between activation levels in BA 44 and combined digit span (r = −0.22, p > 0.3).

3.3 Fractional anisotropy results

The correlation analysis on the FA values inside the thresholded and skeletonized probability map as determined by our combination of TBSS and a deterministic fiber tracking procedure identified different parts of the AF/SLF, which connect the posterior temporal cortex and the IFG, as differentially correlating with the factors ordering and storage: A direct comparison of the correlation with storage as compared to ordering yielded a series of supra-threshold cluster in the middle and posterior portion of the AF/SLF. A direct comparison of the correlation with ordering as compared to storage yielded a small supra-threshold cluster in the frontal portion of the AF/SLF, adjacent to the left IFG. As can be seen in Figure 5, the posterior part of the FA–functional correlation along the AF/SLF is adjacent to the functional cluster in the left PT that was obtained for the main effect of storage and terminates below the middle superior temporal gyrus (STG). Table 2 provides an overview of the significant clusters.

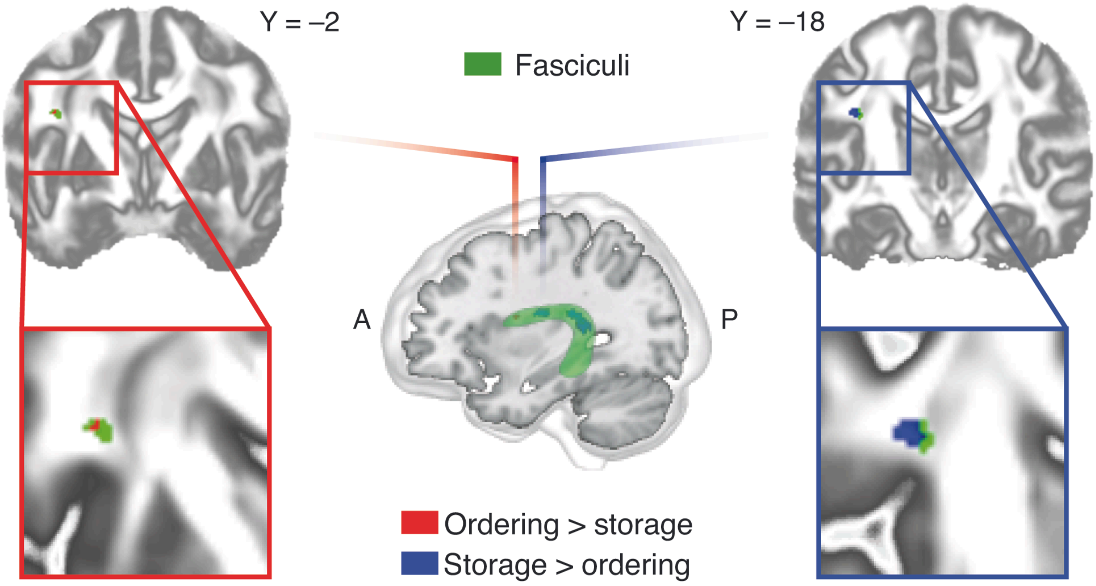

Figure 5: Correlation of FA values with signal change of the storage effect in left TP regions as compared to the ordering effect in the IFG (blue). The regression analysis resulting from the averaging of the individual fiber tracts was carried out in a volume of 1561 voxels of the AF/SLF (green). Clusters were thresholded at $p < 0.05$ and a minimum cluster size of 7 suprathreshold voxels to control for family-wise error at $p < 0.005$.

4 Discussion

Our study set out to separate the neural correlates of ordering and storage during sentence processing in the light of previous evidence on the neuropsychological reality of these two concepts. We also sought to elucidate on the ongoing discussion of the role of Broca's area concerning these processes. Before we turn to discussing the behavioural as well as the functional and structural MRI findings, we want to shortly sketch the reasoning behind our experimental paradigm again.

Our paradigm maximized the contrast along an ordering and a storage dimension, operationalizing argument ordering as an executive operation on the arguments stored in working memory (Wingfield and Butterworth, 1984; Baddeley, 2012), balancing morphosyntactic dependencies across conditions. The ordering factor manipulated the relative argument order, while not interfering with the storage factor. We achieved this by keeping both arguments immediately adjacent to the sentence-final main verb in the short conditions, while keeping either subject or object at the identical, sentence-initial position in the long conditions. The sentence-initial, pre-auxiliar position was chosen to maximize the contrast along the storage dimension—relying on a body of early behavioral (for review, see Nicol and Swinney, 1989) and ERP work (e.g., Kluender and Kutas, 1993; Phillips, 2005) that has consistently shown that arguments are stored across auxiliaries until the main verb is encountered. While the auxiliary provides information on gender and number of the subject argument across conditions, the use of the analytic past perfect tense form (Klein, 2000) made the information about the number and type of arguments available only at the sentence-final man verb (Binder, 2001; Comrie, 1993), balancing storage demands across both short and both long conditions, respectively.

4.1 Behavioral data

Although d'-scores were significantly above chance, for 7 participants, they were relatively low. We assume that the infrequent occurrence of task trials (16.7 % of trials) and the short time window in which participants had to respond (1500 ms) are responsible for the comparably poor performance. In addition to the fact that the inclusion of d'-scores as a regressor of no interest in the fMRI analysis accounted for performance-related variance in the data, it is noteworthy that an exploratory exclusion of these participants from the fMRI analysis affected the statistical results only quantitatively, that is, the overall pattern of activations was not changed—given the low statistical power resulting from the small remaining group size, both effects were robust, with the IFG effect of 179 voxels (z = 3.34, peak at x = –57, y = 17, z = 4), and the TP region effect encompassing multiple temporo-parietal sub-peaks of 52 voxels altogether (z = 2.71, main peaks at x = –60, y = –25, z = 46 and x = –45, y = –28, z = 22).

This is in line with results by Caplan et al. (2008) and Newman et al. (2009), who observed that parts of the left IFG activate more for increased ordering demands, regardless of the experimental task performed or task performance. Furthermore, this is in line with the finding that none of the functional effects in our study showed a correlation with d'-scores. Thus, we are confident that our task did keep participants' attention directed towards the sentences, while leaving the functional results unbiased.

4.2 Functional MRI results

The findings from our functional experiment are straightforward: we found a clear activation for the ordering factor in the left IFG (object-first sentences leading to stronger activation than subject-first sentences) and an as-clear but remote activation for the storage factor (long argument–verb distances leading to stronger activation than short argument-verb distances) in the left TP region. There was no significant interaction between the ordering and storage factor, even though visual inspection of the signal change in the IFG and TP may suggest differently (see Figure 3A and 3C). Moreover, a correlation of memory scores with the storage-related activation of the TP region, but not the IFG region, was observed (see Figure 4).

These findings suggest that the processing of complex sentences relies both on storage of arguments in working memory, supported by left TP regions, and the ordering of arguments, which is supported by Broca's area. The correlation between activity in left TP regions and individual memory capacity further implies that the present “storage” manipulation did indeed tap storage. This highlights the role of the left TP region as a neural substrate of memory storage (e.g., Buchsbaum and D’Esposito, 2008; Jacquemot and Scott, 2006), but more importantly, it emphasises that TP serves this function also in the processing of complex sentences.

The view that Broca’s area necessarily supports sentence processing by its role in subvocal rehearsal is not supported by our data: In the conception of Baddeley (2009; 2012), working memory necessarily involves both storage and rehearsal, the latter providing a constant refreshing of the content stored by the former. Paulesu et al.’s (1993) and subsequent working memory studies (Awh et al., 1996; Awh et al., 1995; Petrides et al., 1993) have conceptualized the neural interplay of storage and rehearsal as a fronto-temporal network involving the IFG (implied in rehearsal) and TP regions (implied in storage). In our study, the IFG did neither exhibit sensitivity to argument–verb distance (the working memory factor in our paradigm), nor did its activation correlate with behavioral measures of working memory ability, unlike TP region activation. To put it differently: while during verbal working memory the IFG may underlie subvocal rehearsal in the phonological loop, it is doubtful that IFG sensitivity to an argument order manipulation in the present study reflects subvocal rehearsal. We will now discuss these findings in more detail.

4.2.1 The left IFG activates for ordering

The increased activity in BA 44 elicited by increased ordering demands (i.e., object-first compared to subject-first sentences) in the present study supports previous cross-linguistic functional neuroimaging research from languages as diverse as German (Friederici et al., 2006b; Obleser et al., 2011; Röder et al., 2002), Japanese (Kim et al., 2009; Kinno et al., 2008) and Hebrew (Ben-Shachar et al., 2003). In all of these studies, increased activation for object- as compared to subject-first sentences was elicited in the left IFG, but not in the left TP region.

First, in German, three studies directly contrasting object- and subject-first sentences reported the inferior pars opercularis (BA 44) to be increasingly active as a function of increased ordering demands. Friederici et al.'s (2006b) study which visually presented sentences, keeping the argument–verb distance constant across conditions, excluded a possible explanation in terms of storage demands. Obleser et al. (2011) used acoustic versions of the same stimuli, again finding activity in BA 44. Both results are in line with Röder et al.'s (2002) finding that object-first sentences in acoustically presented sentences elicit activity in the left IFG.

Second, similar evidence from Japanese strengthens our interpretation. Both Kinno et al. (2008) and Kim et al. (2009) contrasted Japanese object- and subject-first sentences in visual fMRI studies. The fact that Japanese allows for a constant argument–verb distance across both object- and subject-first sentences allowed their experimental paradigms to avoid differences in storage demands across experimental conditions. Both studies found very close areas in the left IFG to increase in activation for non-canonical as compared to canonical sentences.

Third, evidence from a Hebrew study (Ben-Shachar et al., 2003) expands the cross-linguistic picture. Again, Hebrew is a language that allows the fixing of storage while manipulating ordering demands. Ben-Shachar et al.'s (2003) study used auditory presentation of object- or subject-first sentences and obtained activation peaks for the contrast in the left IFG in both a whole-brain and a subsequent ROI analysis.

The fact that our results converge on this body of work strengthens the position that BA 44 as part of Broca's area is engaged in the ordering of arguments; it furthermore suggests that this function is independent of a particular language and input modality. This is further corroborated by English-language studies contrasting the processing of object- and subject-first sentences across input modalities (Constable et al., 2004) and sentences of varying processing difficulty across input modalities (Braze et al., 2011). An overview of these converging results is given in Figure 6.

In spite of this convergence, English-language data have resulted in the claim that the contribution of Broca's area to sentence processing is that of subvocal rehearsal as part of the working memory network (Just et al., 1996; Rogalsky et al., 2008). Unfortunately, the data on which this view is based are not unequivocal, since the relevant studies on ordering and storage in English contrasted subject- and object-relative clauses. While such English clauses certainly scrutinize some aspect of working memory (by varying the number of phrases between an argument and the verb), they also colaterally introduce an ordering manipulation: While the former (short vs. long argument–verb distance) would tax brain regions that subserve storage, the latter (object- vs. subject-first sentences) would rather tax brain regions that are concerned with the ordering of arguments. Thus, the English results are partially ambiguous in that they might be ascribed to either of these factors.

Additionally, clinical evidence demonstrates that the ability to process sentences with increased ordering demands can be independently impaired from subvocal rehearsal abilities (Caplan and Waters, 1999; Waters and Caplan, 1996), and imaging data suggest that subvocal rehearsal during sentence processing does not further increase brain activation in Broca's area (Caplan et al., 2000). The results of our correlation analyses are in line with these reports: If subvocal rehearsal were the driving force behind Broca's area's activity in our study, we would expect activation in BA 44 to correlate with participants' combined digit span scores—instead, only brain activity in the TP region was correlated with combined digit span scores.

In contrast to the above evidence that rehearsal and ordering are conceptually and neurally distinct, Rogalsky et al. (2008) report a decline in sentence processing performance under conditions of articulatory rehearsal. This result, however, was not fully conclusive, in that a control condition (finger tapping during a sentence processing task) also led to a selective decline in sentence processing performance, suggesting that the presence of a secondary task can affect sentence processing behaviorally.

Figure 6 puts these data into a broader perspective, taking into consideration findings from studies inside and outside the sentence processing domain. From this overview figure illustrating the relevant neuroimaging literature it becomes obvious that there is a relative proximity, but no full neuroanatomical overlap between those prefrontal brain regions that support ordering and those regions that support articulatory rehearsal. Thus, there may be a possible differentiation within the prefrontal cortex with respect to ordering of elements and the rehearsal component of working memory which in turn is clearly separated from the storage component located in more posterior regions.

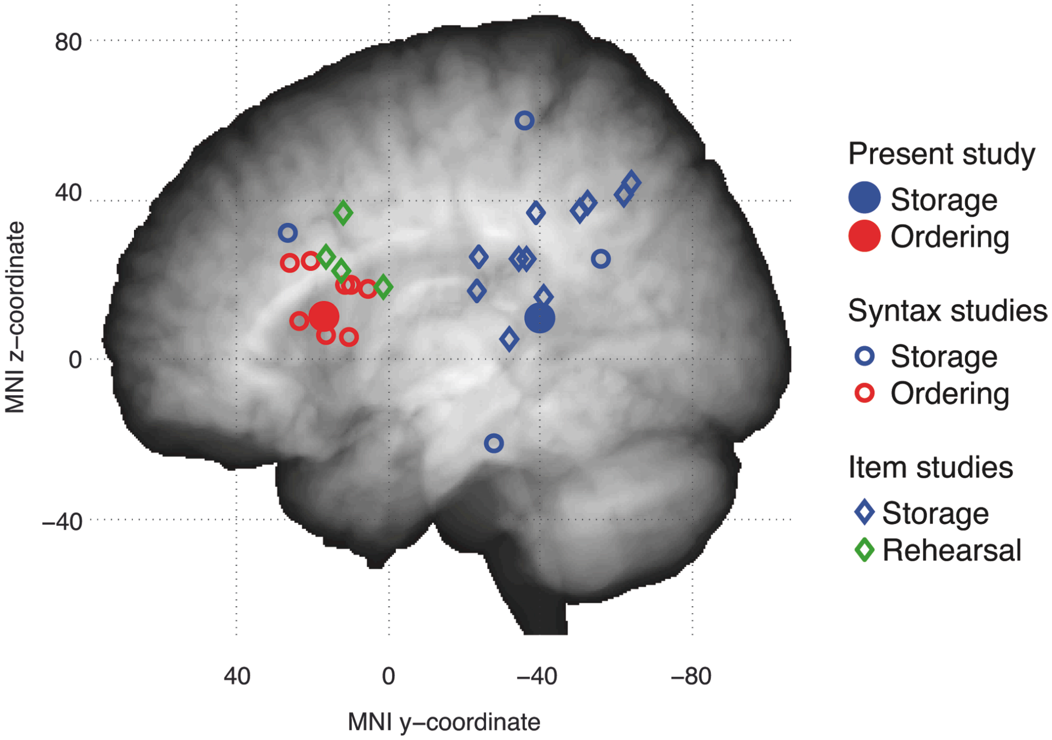

Figure 6: Overview of critical group results from cited studies on verbal WM in sentence-processing and non-sentence-processing paradigms. Circles mark sentence-processing studies, diamonds mark non-sentence-processing studies. It is clearly visible that storage (blue) in both sentence-processing and non-sentence-processing studies is more likely to activate temporo-parietal regions, whereas prefrontal regions show a dissociation between ordering (red circles) and subvocal rehearsal (green diamonds).

Table 2: Overview of significant clusters in the order > storage and storage > order contrast on the FA values inside the skeletonized tract connecting the functional effects (1561 voxels). A 7-voxel threshold at p < 0.05 was applied to achieve whole-tract type I error control at p < 0.005*.

Site	MNI-coordinate			Cluster size (mm3)	Z-score
	X	Y	Z		
Order > storage					
AF/SLF	–38	–2	26	9	2.13
Storage > order					
AF/SLF	–34	–39	21	175	3.56
AF/SLF	–38	–16	28	80	3.33

*Labels according to Mori et al. (2005).

4.2.2 The left TP region activates for storage

This brings us to our second main finding, a cluster in the left TP region that showed both increased activation with increasing storage demands and a correlation with digit span. A number of studies suggest that the left TP region (mainly the PT, but often extending into or peaking in the SMG) is critically involved in storage processes as diverse as storing an ambiguous sentence structure, storing arguments at increasing argument–verb distances and storing non-sentential items in the phonological loop, making left TP regions a strong candidate for the neural substrate of storage, both in sentence processing and item-based tasks (see Figure 6 for an overview of the studies discussed here).

Amongst this work are studies that searched for the neural correlates of the phonological loop (for review see Baddeley et al., 2009), i.e., outside of sentence processing proper. Paulesu et al. (1993) suggested the left SMG as the main region subserving storage. Other imaging studies also consider left posterior regions to subserve storage; as mapped in Figure 6. However, the exact locations of the activation peaks in these studies encompassed various regions, including left superior temporal gyrus (STG) (Kim et al., 2002), SMG (Awh et al., 1996; Clark et al., 2000; Paulesu et al., 1993) and left inferior and posterior parietal cortices (Awh et al., 1995; Becker et al., 1996; Bushara et al., 1999; Clark et al., 2000; Gruber and Von Cramon, 2001, 2003; Jonides et al., 1998; Owen et al., 2005).

For sentence processing, Grossman et al. (2002) found that seniors who show difficulties in processing sentences with increased storage demands activated left parietal cortex relatively less as compared to younger participants. Accordingly, the authors suggest reduced storage resources in seniors amongst the sources of their sentence processing difficulties. The location of the reported regions (see Figure 6) is in line with our findings that left TP activation correlates with digit span. The fact that this correlation is negative fits this interpretation: In our study, participants with relatively better storage abilities (as tested by digit span) showed relatively less activation in the left TP region, suggesting more efficient storage in these participants.

Additional evidence for a general storage component in left posterior cortex comes from clinical studies on patients with damage to the left STG, SMG, or TP region. The observation that damage to the STG causes impaired storage abilities has been made by Leff et al. (2009), who found gray-matter integrity in the STG in stroke patients to correlate with digit span. Recently, these findings have been augmented by studies on conduction aphasia (Buchsbaum et al., 2011; Fridriksson et al., 2010) demonstrating a relation between speech repetition problems, phonological working memory deficits and gray-matter damage to the left SMG and TP region. While such results can be ambiguous in that conduction aphasics usually suffer from both gray- and white-matter lesions, direct causal evidence is provided by a repetitive transcranial magnetic stimulation (rTMS) study (Romero et al., 2006) showing that rTMS applied to the SMG causes a decrease in digit span.

These somewhat heterogenous localizations in the posterior regions across studies may result from distinct working memory sub-processes tapped by the various paradigms employed, since in addition to rehearsal and storage as discussed above, retrieval is assumed to be a separate component of working memory. Evidence for a segregation of storage and retrieval comes from a recent working memory fMRI study by Ravizza et al. (2010). In a series of ROI analyses, these researchers found storage of verbal material to correlate with brain activity in the left posterior STG, whereas retrieval of verbal material activated the left temporo-parietal junction. Functional heterogeneity in this area related to differential working memory processes has been previously suggested by Henson et al. (1999) and Buchsbaum et al. (2001), who report distinct sub-peaks during a working memory task in the posterior STG and very inferior parietal cortex. Although our results can not decide on the neural underpinnings of separable storage and retrieval sub-processes, the activation we observed in the left TP region involved two sub-peaks (see Table 1), one clearly in the PT, one in the SMG. Hence, we may speculate that our experimental manipulation of argument–verb distance taxed both storage and retrieval, due to the longer retention interval and accordingly higher re-activation demands induced by memory decay (Baddeley et al., 2009).

A final and important issue that needs to be addressed is the apparent contrast between the present data and some previous studies on working memory demands during sentence processing. In English, direct comparisons of either pronoun binding and argument–verb dependencies (Santi and Grodzinsky, 2007) or argument–verb dependencies and embedded sentences (Santi and Grodzinsky, 2010) yielded brain activation in BA 45. Similarly, a German study found activity related to subject argument–verb distance in the inferior frontal sulcus, dorsal to and extending into BA 45 (Makuuchi, et al. 2009) whereby the argument–verb distance manipulation introduced a difference in the number of argument–verb dependencies. Finally, a second German study found BA 45 activation for pronoun–verb distance, but only for the comparison between object- and subject-first sentences (Fiebach et al., 2005). We suggest that the comparison of different syntactic dependencies across conditions in the above studies may have reflected the engagement of a syntactic working memory system, which has been proposed to be distinct from the working memory used in other verbal tasks, and found to activate BA 45 (Caplan et al., 2000; Lewis et al., 2006; Van Dyke, 2007). In contrast to the above studies, our paradigm kept the type of syntactic dependency constant across conditions and required the storage of a given noun phrase across a variable distance. Such an interpretation is corroborated by the possibility that our experimental task was solveable using phonological strategies, which is also true for the digit span task used in our correlation analysis. Thus, we suggest that the contrast between temporo-parietal and inferior frontal brain activations reflects the difference between phonological and syntactic working memory, respectively. Still, this suggestion can only count as a hypothesis for future research.

4.3 Diffusion MRI results

Deterministic fiber tracking between the observed functional effects suggested a dorsal connection, including the arcuate fascicle and the superior longitudinal fascile

(AF/SLF). As to the correlation analysis of the FA values in the AF/SLF with our functional effects, a series of clusters was obtained. Clusters of FA along the superior and posterior parts of the AF/SLF correlated with increased storage demands as compared to increased ordering demands, and a single cluster in the more anterior part correlated with increasing ordering as compared to storage demands. The involvement of the AF/SLF, but not the inferior longitudinal fasciculus, suggests that our effects are related to the dorsal pathway, linking the auditory cortex to the IFG via TP regions (Catani and Jones, 2005; Weiller et al., 2009). Generally, both the effects of ordering as compared to storage and storage as compared to ordering along the left AF/SLF are in accordance with the view that this tract is involved in sentence processing (Friederici, 2009). This is in line with probabilistic fiber tracking results, which were based on functionally defined seeding points in the IFG (pars opercularis / BA 44; Friederici et al., 2006b) and are similar to our data with regard to this. Specifically, it seems anatomically plausible that the neural basis of sentence processing involves an ordering process as subserved by the IFG that queries a storage component in TP regions through the AF/SLF.

The literature on an involvement of different fiber tracts in working memory from healthy populations is sparse, although initial DTI work (Charlton et al., 2010) is available. Charlton et al. (2010) used a composite working memory score collating different working memory measures to explore the white-matter network underlying working memory. Their results show a distributed array of mean diffusivity and FA clusters in both hemispheres, amongst which is a tract that connects BA 40 to inferior frontal areas. But since none of the measures employed by Charlton et al. (2010) may only have tapped storage, their results do not easily map directly onto our sentence processing data. Thus, to our knowledge, our study is the first to show that the left

AF/SLF is relevant to more specific sub-tasks of sentence processing, that is, on the one hand storing and holding an argument in phonological working memory over a certain argument–verb distance of the sentence and on the other hand manipulating the argument order by syntactic working memory such that the who and the whom in the sentence are not confused. In combination with Charlton et al.'s (2010) results, the present data may suggest a common role of the AF/SLF in storage both in- and outside the sentence processing domain.

The common role of the AF/SLF may best be described as mediating a storage component in TP regions to a rehearsal component located in the dorsal prefrontal region for tasks outside of sentence processing and to an ordering system located in Broca's area for sentence processing proper. This interpretation fits both our finding of a correlation with storage involved in sentence processing (as measured by brain activation during increasing argument–verb distances) and outside of sentence processing (as measured by digit span). Also, the relatively weaker and less extensive correlations with argument order in the anterior portion of the AF/SLF as opposed to rather extensive effects of argument–verb distance suggest that the functional role of this tract is less specific to ordering itself than it is general to the storage of processing-relevant information.

In addition to work on healthy participants, there is evidence on the role of the left AF/SLF in sentence processing from patients with primary progressive aphasia (Wilson et al., 2011) and patients with conduction aphasia, who also show reduced working memory capacities (Buchsbaum et al., 2011). Investigating patients with primary progressive aphasia, Wilson et al. (2011) found that damage to the AF/SLF causes severe sentence processing difficulties in primary progressive aphasics, which was not the case for patients with damage to the ventrally located extreme capsule fiber system and uncinate fasciculus. The studies on conduction aphasia are of different degrees of neuroanatomical specificity. Friedmann and Gvion (2003) report that conduction aphasics show both reduced digit span and particular problems in argument retrieval at long argument–verb distances. Since the authors are not precise about lesion site and extent, this study does not allow to draw strong neuroanatomical conclusions. Yamada et al. (2007), however, report a case in which selective damage to the posterior part of the AF/SLF resulted in conduction aphasia, sparing the gray matter in the left TP region. Finally, Fridriksson et al. (2010) found that impaired speech repetition in conduction aphasics co-occurs with both damage to the white matter underlying the left SMG and reduced gray-matter integrity in the left SMG and temporo-parietal junction, converging on data by Baldo et al. (2008). Our data agree with both the gray-matter and white-matter proposal. Together, this is solid ground for the proposal that the integrity of the left AF/SLF is crucial to the storage component of working memory and its linkage to Broca's area during the processing of argument–verb dependencies.

Turning towards methodology, it is worth to briefly discuss our white-matter analysis—in particular because we used the fiber tracking results not as a dependent variable in the correlation analysis, but merely as the search volume for the regression of functional data across FA-maps. An alternative approach would have been to compare the reconstructed individual tracts themselves, e.g. the number of streamlines. However, Wakana et al. (2007) and Heiervang et al. (2006) have pointed out that there is no proportional relation between individual FA values and the number of streamlines extracted by fiber tracking, and even less so at small sample sizes. Fiber tracking is sensitive to anatomical variability along the to-be-reconstructed tract, such as overall tract length, crossing fibers or tract diameter, inter alia. Thus, while fiber tracking is a valuable tool for anatomical description, FA values are a functionally more relevant and statistically more reliable measure of local microstructural properties (Reich et al., 2010).

Another issue with this FA analysis might be that we used our functional effects both as seeding points in the fiber tracking analysis and as regressors in the correlation analysis between FA values and differences in functional activation. In principle, Kriegeskorte et al. (2009) point out that such an analysis could yield invalid results due to an overlap between dependent and independent variables. However, even though we used our functional activation clusters as gray-matter seeding points, the DTI procedure reconstructed only white-matter tracts. In addition, these tracts were masked with the TBSS skeleton, which also only contains areas of high FA across participants, that is, white matter. Since the selection data at this point were not the functional data themselves, but a reconstructed tract volume, we consider the dependent measure (i.e. voxel-wise FA of the fiber tracts connecting the IFG and TP region) to be different from the independent measure (i.e. the individual difference in functional activation strength).

5. Conclusion

Our results show that, in sentence processing, the storage of arguments over increased argument–verb distance and the ordering of arguments rely on distinct neural subsystems. The direct comparsion of these factors within one stimulus set shows that Broca’s area (IFG) is mainly concerned with ordering, and less so with storage. Storage of a single argument in working memory across a given argument–verb distance activates the left TP region; a region classically assumed to subserve working memory storage. The data, moreover, provide direct evidence for the role of the left

AF/SLF in interfacing working memory storage and ordering. The results suggest that a minimal cognitive architecture of sentence processing can be rooted in the interplay of concrete cognitive concepts, such as ordering and storage.

Acknowledgments

This research was supported by the Max Planck Society. The authors are grateful to Celia Sommer, Simone Wipper, Annett Wiedemann and Anke Kummer, who helped acquire the MR Data. We also thank two anonymous reviewers for their detailed and constructive suggestions.

References

Amici, S., Brambati, S.M., Wilkins, D.P., Ogar, J., Dronkers, N.L., Miller, B.L., Gorno-Tempini, M.L., 2007. Anatomical correlates of sentence comprehension and verbal working memory in neurodegenerative disease. Journal of Neuroscience 27, 6282–6290.

Andrade, A., Paradis, A.L., Rouquette, S., Poline, J.B., 1999. Ambiguous results in functional neuroimaging data analysis due to covariate correlation. NeuroImage 10, 483–486.

Anwander, A., Tittgemeyer, M., Von Cramon, D.Y., Friederici, A.D., Knösche, T.R., 2007. Connectivity-based parcellation of Broca's area. Cerebral Cortex 17, 816–825.

Awh, E., Jonides, J., Smith, E.E., Schumacher, E.H., Koeppe, R.A., Katz, S., 1996. Dissociation of storage and rehearsal in working memory: PET evidence. Psychological Science 7, 25–31.

Awh, E., Smith, E.E., Jonides, J., 1995. Human rehearsal processes and the frontal lobes: PET evidence. Annals of the New York Academy of Science 769, 97–118.

Baayen, H., Gulikers, L., Piepenbrock, R., 1995. The CELEX lexical database. Linguistic Data Consortium, Philadelphia, PA.

Baddeley, A., Eysenck, M.W., Anderson, M.C., 2009. Memory. Psychology Press, Hove.

Baddeley, A., 2012. Working memory: theories, models, and controversies. Annual review of psychology 63, 1–29.

Baldo, J.V., Klostermann, E.C., Dronkers, N.F., 2008. It's either a cook or a baker: Patients with conduction aphasia get the gist but lose the trace. Brain and Language 105, 134–140.

Basser, P.J., Mattiello, J., Lebihan, D., 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. Journal of Magnetic Resonance Imaging Series B 103, 247–247.

Basser, P.J., Pierpaoli, C., 1996. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. Journal of Magnetic Resonance Imaging Series B 111, 209–219.

Becker, J.T., Mintun, M., Aleva, K., Wiseman, M., Nichols, T., DeKosky, S., 1996. Compensatory reallocation of brain resources supporting verbal episodic memory in Alzheimer's disease. Neurology 46, 692.

Ben-Shachar, M., Hendler, T., Kahn, I., Ben-Bashat, D., Grodzinsky, Y., 2003. The neural reality of syntactic transformations. Psychological Science 14, 433–440.

Biemann, C., Bordag, S., Quasthoff, U., Wolff, C., 2004. Web services for language resources and language technology applications. Proceedings of the Fourth International Conference on Language Resources and Evaluation, Lissabon, Portugal.

Boersma, P., Weenink, D., 2001. Praat, a system for doing phonetics by computer. Glot International 5, 341–345.

Binder, K.S., Duffy, S.A., and Rayner, K. (2001). The Effects of Thematic Fit and Discourse Context on Syntactic Ambiguity Resolution. Journal of Memory and Language 44, 297–324.

Bornkessel, I., Zysset, S., Friederici, A.D., von Cramon, D.Y., Schlesewsky, M. Who did what to whom? The neural basis of argument hierarchies during language comprehension. NeuroImage 26 (1): 221–233.

Brauer, J., Anwander, A., Friederici, A.D., 2011. Neuroanatomical prerequisites for language functions in the maturing brain. Cerebral Cortex 21, 459–466.

Braze, D., Mencl, W.E., Tabor, W., Pugh, K.R., Todd Constable, R., Fulbright, R.K., Magnuson, J.S., Van Dyke, J.A., Shankweiler, D.P., 2011. Unification of sentence processing via ear and eye: An fMRI study. Cortex 47, 416–431.

Brett, M., Anton, J., Valabregue, R., Poline, J., 2002. Region of interest analysis using an SPM toolbox. NeuroImage 16, 1140–1141.

Buchsbaum, B.R., Hickok, G., Humphries, C., 2001. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. Cognitive Science 25, 663–678.

Buchsbaum, B.R., Baldo, J., Okada, K., Berman, K.F., Dronkers, N., D'Esposito, M., Hickok, G., 2011. Conduction aphasia, sensory-motor integration, and phonological short-term memory - an aggregate analysis of lesion and fMRI data. Brain and Language 119, 119–128.

Buchsbaum, B.R., D'Esposito, M., 2008. The search for the phonological store: from loop to convolution. Journal of Cognitive Neuroscience 20, 762–778.

Bushara, K.O., Weeks, R.A., Ishii, K., Catalan, M.J., Tian, B., Rauschecker, J.P., Hallett, M., 1999. Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. Nature Neuroscience 2, 759–766.

Caplan, D., Alpert, N., Waters, G., Olivieri, A., 2000. Activation of Broca’s Area by Syntactic Processing Under Conditions of Concurrent Articulation. Human Brain Mapping 9, 65–71.

Caplan, D., Chen, E., Waters, G., 2008. Task-dependent and task-independent neurovascular responses to syntactic processing. Cortex 44, 257–275.

Caplan, D., Waters, G.S., 1990. Short-term memory and language comprehension: A critical review of the neuropsychological literature. In: Vallar, G., Shallice, T. (Eds.), Neuropsychological impairments of short-term memory. Cambridge University Press, Cambridge, pp. 337–389.

Caplan, D., Waters, G.S., 1999. Verbal working memory and sentence comprehension. Behavioral Brain Science 22, 77–94.

Catani, M., Jones, D.K., 2005. Perisylvian language networks of the human brain. Annals of Neurology 57, 8–16.

Charlton, R.A., Barrick, T.R., Lawes, I., Markus, H.S., Morris, R.G., 2010. White matter pathways associated with working memory in normal aging. Cortex 46, 474–489.

Clahsen, H., Featherston, S., 1999. Antecedent Priming at Trace Positions: Evidence from German Scrambling. Journal of Psycholinguistic Research 28, 415–437.

Clark, C.R., Egan, G.F., McFarlane, A.C., Morris, P., Weber, D., Sonkkilla, C., Marcina, J., Tochon-Danguy, H.J., 2000. Updating working memory for words: A PET activation study. Human Brain Mapping 9, 42–54.

Comrie, B., 1993. Argument structure. In: Jacobs, J., von Stechow, A., Sternefeld, W., Vennemann, T. (Eds.), Syntax. Ein internationales Handbuch zeitgenˆssischer Forschung. de Gruyter, Berlin, pp. 905–914.Constable, R.T., Pugh, K.R., Berroya, E., Mencl, W.E., Westerveld, M., Ni, W., Shankweiler, D., 2004. Sentence complexity and input modality effects in sentence comprehension: an fMRI study. NeuroImage 22, 11–21.

Cooke, A., Zurif, E.B., DeVita, C., Alsop, D., Koenig, P., Detre, J., Gee, J., Pinango, M., Balogh, J., Grossman, M., 2002. Neural basis for sentence comprehension: Grammatical and short-term memory components. Human Brain Mapping 15, 80–94.

Daneman, M., Carpenter, P.A., 1980. Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior 19, 450–466.

D'Esposito, M., Postle, B., Ballard, D., Lease, J., 1999. Maintenance versus Manipulation of Information Held in Working Memory: An Event-Related fMRI Study. Brain and Cognition 41, 66–86.

Eickhoff, S., Stephan, K., Mohlberg, H., Grefkes, C., Fink, G., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage 25, 1325–1335.

Felser, C., Clahsen, H., Münte, T.F., 2003. Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. Brain and Language 87, 345–354.

Ferreira, F., Henderson, J.M., 1993. Reading processes during syntactic analysis and reanalysis. Canadian Journal of Experimental Psychology 47, 247–247.

Fiebach, C.J., Schlesewsky, M., Friederici, A.D., 2001. Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. Journal of Psycholinguistic Research 30, 321–338.

Fiebach, C.J., Schlesewsky, M., Lohmann, G., von Cramon, D.Y., Friederici, A.D., 2005. Revisiting the role of Broca's area in sentence processing: Syntactic integration versus syntactic working memory. Human Brain Mapping 24, 79–91.

Fillard, P., Pennec, X., Arsigny, V., Ayache, N., 2007. Clinical DT-MRI estimation, smoothing, and fiber tracking with log-Euclidean metrics. IEEE Transactions on Medical Imaging 26, 1472–1482.

Fridriksson, J., Kjartansson, O., Morgan, P.S., Hjaltason, H., Magnusdottir, S., Bonilha, L., Rorden, C., 2010. Impaired speech repetition and left parietal lobe damage. Journal of Neuroscience 30, 11057–11061.

Friederici A.D., 2002. Towards a neural basis of auditory sentence processing. Trends in Cognitive Science 6, 78–84.

Friederici, A.D., 2009. Pathways to language: fiber tracts in the human brain. Trends in Cognitive Science 13, 175–181.

Friederici, A.D, Opitz, B., von Cramon, D.Y., 2000. Segregating semantic and syntactic aspects of processing in the human brain: An fMRI investigation of different word types. Cerebral Cortex 10, 698–705.

Friederici, A.D., Bahlmann, J., Heim, S., Schubotz, R.I., Anwander, A., 2006a. The brain differentiates human and non-human grammars: functional localization and structural connectivity. Proceedings of the National Academy of Sciences of the USA 103, 2458–2463.

Friederici, A.D., Fiebach, C.J., Schlesewsky, M., Bornkessel, I.D., von Cramon, D.Y., 2006b. Processing Linguistic Complexity and Grammaticality in the Left Frontal Cortex. Cerebral Cortex 16, 1709–1717.

Friedman, D., Simson, R., Ritter, W., Rapin, I., 1975. The late positive component (P300) and information processing in sentences. Electroencephalography and clinical neurophysiology 38, 255–262.

Friedmann, N., Gvion, A., 2003. Sentence comprehension and working memory limitation in aphasia: A dissociation between semantic-syntactic and phonological reactivation. Brain and Language 86, 23–39.

Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A., 2002. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magnetic Resonance in Medicine 47, 1202–1210.

Grossman, M., Cooke, A., DeVita, C., Alsop, D., Detre, J., Chen, W., Gee, J., 2002. Age-related changes in working memory during sentence comprehension: an fMRI study. NeuroImage 15, 302–317.

Gruber, O., Von Cramon, D.Y., 2001. Domain-specific distribution of working memory processes along human prefrontal and parietal cortices: a functional magnetic resonance imaging study. Neuroscience Letters 297, 29–32.

Gruber, O., Von Cramon, D.Y., 2003. The functional neuroanatomy of human working memory revisited:: Evidence from 3-T fMRI studies using classical domain-specific interference tasks. NeuroImage 19, 797–809.

Heiervang, E., Behrens, T., Mackay, C., Robson, M., Johansen-Berg, H., 2006. Between session reproducibility and between subject variability of diffusion MR and tractography measures. NeuroImage 33, 867–877.

Henson, R.N., Burgess, N., Frith, C.D., 1999. Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. Neuropsychologia 38, 426–440.

Jacquemot, C., Scott, S.K., 2006. What is the relationship between phonological short-term memory and speech processing? Trends in Cognitive Sciences 10, 480–486.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17, 825–841.

Jonides, J., Schumacher, E.H., Smith, E.E., Koeppe, R.A., Awh, E., Reuter-Lorenz, P.A., Marshuetz, C., Willis, C.R., 1998. The role of parietal cortex in verbal working memory. Journal of Neuroscience 18, 5026–5034.

Just, M.A., Carpenter, P.A., Keller, T.A., Eddy, W.F., Thulborn, K.R., 1996. Brain activation modulated by sentence comprehension. Science 274, 114–116.

Kim, J., Kim, M.S., Lee, J.S., Lee, D.S., Lee, M.C., Kwon, J.S., 2002. Dissociation of working memory processing associated with native and second languages: PET investigation. NeuroImage 15, 879–891.

Kim, J., Koizumi, M., Ikuta, N., Fukumitsu, Y., Kimura, N., Iwata, K., Watanabe, J., Yokoyama, S., Sato, S., Horie, K., others, 2009. Scrambling effects on the processing of Japanese sentences: An fMRI study. Journal of Neurolinguistics 22, 151–166.

Kinno, R., Kawamura, M., Shioda, S., Sakai, K.L., 2008. Neural correlates of noncanonical syntactic processing revealed by a picture sentence matching task. Human Brain Mapping 29, 1015–1027.

Klein, W. (2000). An Analysis of the German Perfekt. Language 76, 358–382.

Kluender, R., Kutas, M., 1993. Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. Journal of Cognitive Neuroscience 5, 196–214.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.I., 2009. Circular analysis in systems neuroscience – the dangers of double dipping. Nature Neuroscience 12, 535–540.

Lazar, M., Weinstein, D.M., Tsuruda, J.S., Hasan, K.M., Arfanakis, K., Meyerand, M.E., Badie, B., Rowley, H.A., Haughton, V., Field, A., 2003. White matter tractography using diffusion tensor deflection. Human Brain Mapping 18, 306–321.

Leff, A.P., Schofield, T.M., Crinion, J.T., Seghier, M.L., Grogan, A., Green, D.W., Price, C.J., 2009. The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: evidence from 210 patients with stroke. Brain, 3401–3410.

Lewis, R.L., 1996. Interference in short-term memory: The magical number two (or three) in sentence processing. Journal of Psycholinguistic Research 25, 93–115.

Lu, L., Leonard, C., Thompson, P., Kan, E., Jolley, J., Welcome, S., Toga, A., Sowell, E., 2007. Normal developmental changes in inferior frontal gray matter are associated with improvement in phonological processing: a longitudinal MRI analysis. Cerebral Cortex 17, 1092–1099.

Macmillan, N.A., Creelman, C.D., 2005. Detection Theory: A User's Guide, 2 ed. Lawrence Erlbaum Associates, Mahwan, NJ.

Makuuchi, M., Bahlmann, J., Anwander, A., Friederici, A.D., 2009. Segregating the core computational faculty of human language from working memory. Proceedings of the National Academy of Sciences of the USA 106, 8362–8367.

Mori, S., Wakana, S., Nagae-Poetscher, L., Van Zijl, P, 2006. MRI atlas of human white matter. Elsevier, Amsterdam.

Mugler III, J.P., Brookeman, J.R., 1990. Three dimensional magnetization prepared rapid gradient echo imaging (3D MP RAGE). Magnetic Resonance in Medicine 15, 152–157.

Nakano, Y., Felser, C., Clahsen, H., 2002. Antecedent priming at trace positions in Japanese long-distance scrambling. Journal of Psycholinguistic Research 31, 531–571.

Newman, S.D., Lee, D., Ratliff, K.L., 2009. Off-line sentence processing: What is involved in answering a comprehension probe? Human Brain Mapping 30, 2499–2511.

Nicol, J.L., Fodor, J.D., Swinney, D., 1994. Using cross-modal lexical decision tasks to investigate sentence processing. Journal of Experimental Psychology: Learning, Memory, and Cognition 20, 1229–1238.

Nicol, J.L., Swinney, D., 1989. The role of structure in coreference assignment during sentence comprehension. Journal of Psycholinguistic Research 18, 5–19.

Novais-Santos, S., Gee, J., Shah, M., Troiani, V., Work, M., Grossman, M., 2007. Resolving sentence ambiguity with planning and working memory resources: Evidence from fMRI. NeuroImage 37, 361–378.

Obleser, J., Meyer, L., Friederici, A.D., 2011. Dynamic assignment of neural resources in auditory comprehension of complex sentences. NeuroImage 4, 2310–2310.

Oldfield, R., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E., 2005. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. Human Brain Mapping 25, 46–59.

Paulesu, E., Frith, C., Frackowiak, R., 1993. The neural correlates of the verbal component of working memory. Nature 362, 342–345.

Petrides, M., Alivisatos, B., Meyer, E., Evans, A.C., 1993. Functional activation of the human frontal cortex during the performance of verbal working memory tasks. Proceedings of the National Academy of Sciences of the USA 90, 878–882.

Phillips, C., Kazanina, N., Abada, S.H., 2005. ERP effects of the processing of syntactic long-distance dependencies. Cognitive Brain Res 22, 407–428.

Ravizza, S.M., Hazeltine, E., Ruiz, S., Zhu, D.C., 2011. Left TPJ activity in verbal working memory: Implications for storage- and sensory-specific models of short term memory. NeuroImage 55, 1836–1846.

Reese, T., Heid, O., Weisskoff, R., Wedeen, V., 2003. Reduction of eddy current induced distortion in diffusion MRI using a twice refocused spin echo. Magnetic Resonance in Medicine 49, 177–182.

Reich, D.S., Ozturk, A., Calabresi, P.A., Mori, S., 2010. Automated vs. conventional tractography in multiple sclerosis: Variability and correlation with disability. NeuroImage 49, 3047-3056.

Röder, B., Stock, O., Neville, H., Bien, S., Rösler, F., 2002. Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: a functional magnetic resonance imaging study. NeuroImage 15, 1003–1014.

Rogalsky, C., Hickok, G., 2010. The Role of Broca's Area in Sentence Comprehension. Journal of Cognitive Neuroscience, 1–17.

Rogalsky, C., Matchin, W., Hickok, G., 2008. Broca's area, sentence comprehension, and working memory: an fMRI study. Frontiers in Human Neuroscience 2, 1–13.

Romero, L., Walsh, V., Papagno, C., 2006. The neural correlates of phonological short-term memory: a repetitive transcranial magnetic stimulation study. Journal of Cognitive Neuroscience 18, 1147–1155.

Santi, A., Grodzinsky, Y., 2007. Working memory and syntax interact in Broca's area. NeuroImage 37, 8–17.

Santi, A., Grodzinsky, Y., 2010. fMRI adaptation dissociates syntactic complexity dimensions. NeuroImage 51, 1285–1293.

Smith, E.E., Jonides, J., 1998. Neuroimaging analyses of human working memory. Proceedings of the National Academy of Sciences of the USA 95, 12061–12068.

Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. NeuroImage 31, 1487–1505.

Tabachnick, B.G., Fidell, L.S., Osterlind, S.J., 2001. Using multivariate statistics. HarperCollins, New York.

Talairach, J., Tournoux, P., 1988. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. Thieme, New York.

Tewes, U., 1994. Hamburg-Wechsler-Intelligenztest für Erwachsene, Revision 1991. Huber, Bern.

Ueno, M., Kluender, R., 2003. Event-related brain indices of Japanese scrambling. Brain and Language 86, 243–271.

Upadhyay, J., Silver, A., Knaus, T.A., Lindgren, K.A., Ducros, M., Kim, D.S., Tager-Flusberg, H., 2008. Effective and structural connectivity in the human auditory cortex. Journal of Neuroscience 28, 3341–3349.

Van Dyke, J.A., 2007. Interference effects from grammatically unavailable constituents during sentence processing. Journal of Experimental Psychology: Learning, Memory and Cognition 33, 407–430.

Van Petten, C.V., 1993. A comparison of lexical and sentence-level context effects in event-related potentials. Language and Cognitive Processes 8, 485–531.

Wager, T.D., Keller, M.C., Lacey, S.C., Jonides, J., 2005. Increased sensitivity in neuroimaging analyses using robust regression. NeuroImage 26, 99–113.

Wakana, S., Caprihan, A., Panzenboeck, M.M., Fallon, J.H., Perry, M., Gollub, R.L., Hua, K., Zhang, J., Jiang, H., Dubey, P., 2007. Reproducibility of quantitative tractography methods applied to cerebral white matter. NeuroImage 36, 630-644.

Waters, G.S., Caplan, D., 1996. The measurement of verbal working memory capacity and its relation to reading comprehension. Quarterly Journal of Experimental Psychology A 49, 51–79.

Weiller, C., Musso, M., Rijntjes, M., Saur, D., 2009. Please don't underestimate the ventral pathway in language. Trends in Cognitive Science 13, 369–370.

Wilson, S.M., Galantucci, S., Tartaglia, M.C., Rising, K., Patterson, D.K., Henry, M.L., Ogar, J.M., DeLeon, J., Miller, B.L., Gorno-Tempini, M.L., 2011. Syntactic processing depends on dorsal language tracts. Neuron 72, 397–403.

Wingfield, A., Butterworth, B., 1984. Running memory for sentences and parts of sentences: Syntactic parsing as a control function in working memory. In: Bouma, H., Bouwhuis, D. (Eds.), Attention and performance X: Control of language processes. Lawrence Erlbaum Associates, London, pp. 351–364.

Yamada, K., Nagakane, Y., Mizuno, T., Hosomi, A., Nakagawa, M., Nishimura, T., 2007. MR tractography depicting damage to the arcuate fasciculus in a patient with conduction aphasia. Neurology 69, 321.