

## Hierarchical artificial grammar processing engages Broca's area

Jörg Bahlmann, Ricarda I. Schubotz, Angela D. Friederici\*

Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

### ARTICLE INFO

#### Article history:

Received 23 November 2007

Revised 18 April 2008

Accepted 26 April 2008

Available online 7 May 2008

#### Keywords:

Artificial grammar

fMRI

Hierarchical processing

Broca's area

Language

Sequence processing

### ABSTRACT

The present fMRI study investigates the neural basis of hierarchical processing using two types of artificial grammars: one governed by rules of adjacent dependencies and the other by rules of hierarchical dependencies. The adjacent dependency sequences followed the rule  $(AB)^n$ , at which simple transitions between two types of syllable categories were generated (e.g.  $A_1B_1A_2B_2$ ). The hierarchical syllable sequences followed the rule  $A^nB^n$ , generating a center-embedded structure (e.g.  $A_2A_1B_1B_2$ ) the learning of which required the processing of hierarchical dependencies. When comparing the processing of hierarchical dependencies to adjacent dependencies, significantly higher activations were observed in Broca's area and the adjacent rim of the ventral premotor cortex (BA 44/6) in addition to some several other cortical and sub-cortical regions. These results indicate that Broca's area is part of a neural circuit that is responsible for the processing of hierarchical structures in an artificial grammar.

© 2008 Elsevier Inc. All rights reserved.

### Introduction

A recent discussion concerning the biology of language has centered around the question of whether the ability to deal with hierarchical structures and recursion is a key feature of human language processing (Hauser et al., 2002; Pinker and Jackendoff, 2005; Perruchet and Rey, 2005; de Vries et al., 2007; Corballis, 2007; Hauser et al., 2007).

Brain imaging data has revealed that processing of hierarchical grammatical structures be they part of a natural or artificial grammar, recruited Broca's area, i.e. Brodmann's Areas 44 and 45 in the left posterior inferior frontal gyrus (Tettamanti et al., 2002; Musso et al., 2003; Friederici et al., 2006a,b) and so does the processing of elements moved out of their original position in a syntactic structure (Ben-Shachar et al., 2003; Santi and Grodzinsky, 2007).

In a recent study on the question of specific status of the ability to process hierarchical structures, humans and monkeys were compared in their ability to process two types of artificial grammar rules, namely hierarchical structures ( $A^nB^n$ ) and probabilistic structures  $(AB)^n$  (Fitch and Hauser, 2004). The rule  $A^nB^n$  allows the generation of center-embeddings and hierarchical dependencies between elements of the category A and category B. In contrast, the rule  $(AB)^n$  determines local transitions between two types of categories (A and B). In this behavioral study human, but not non-human primates were found to be able to learn the  $A^nB^n$  rule.

However, it appears that the processing of the  $A^nB^n$  rule, as applied in the experiment by Fitch and Hauser (2004), does not necessarily rely on the generation of hierarchical representational structures. Perruchet and Rey (2005) claimed that sequences following the rule  $A^nB^n$  does not require the representation of center-embedded structures of corresponding categories, but rather allows for the computation by a more simple strategy such as counting the elements of the respective categories (see also Hauser et al., 2007 for a similar argument). Interestingly, in a recent study (Gentner et al., 2006) it has been shown that song birds (European starlings) are able to process an  $A^nB^n$  rule. This either means that the processing of this rule is not specifically human, or that an alternative, simpler strategy was applied by these animals. In their study counting of A-category items and matching those with the number of B-category items were assumed as a possible strategy.

This same criticism holds for a recent functional magnetic resonance imaging (fMRI) study investigating the neural basis of processing  $(AB)^n$  and  $A^nB^n$  structured sequences (Friederici et al., 2006a). In this study, similar to the prior behavioral studies with animals and humans, there was no particular element of category A that corresponded to a particular element of category B in a given sequence, thus allowing for processing based on counting category membership in principle. Interestingly, however, the fMRI study reported activation in Broca's area for the comparison of incorrect versus correct hierarchical dependency rule, but not for the respective comparison for the adjacent dependency rule. Given prior fMRI research on processing natural language the observed activation in Broca's area was taken to reflect hierarchical processing rather than a counting strategy. But admittedly, a

\* Corresponding author. Max Planck Institute for Human Cognitive and Brain Sciences, PO Box 500 355, 04103 Leipzig, Germany. Fax: +49 3 41 99 40 113.

E-mail address: [angelafr@cbs.mpg.de](mailto:angelafr@cbs.mpg.de) (A.D. Friederici).

counting strategy cannot be excluded for the particular grammar realizations used in this study. Moreover, as no direct comparison of the hierarchical versus adjacent dependency rule was possible due to the fact that the different artificial grammars were learned by different groups of participants, the activation observed in this study may reflect aspects of processing ungrammaticality in the two types of sequences.

In the present fMRI study, a hierarchical  $A^nB^n$  grammar type which necessarily required hierarchical processing was compared to an artificial grammar that generates adjacent dependencies (AB)<sup>n</sup>. Mathematically, the generation of category sequences according to the expression  $A^nB^n$  engages a recursive/hierarchical procedure, which produces center-embeddings. A production rule like  $S \rightarrow ASB$  and  $S \rightarrow AB$  (while  $S$  could be seen as a “sentence”) generates the following examples (1)  $S \rightarrow AB$ ; (2)  $S \rightarrow AAB$ ; and (3)  $S \rightarrow AAABB$ . This routine calls itself recursively.  $A$  and  $B$  are attached adjacent to the  $S$  symbol in the middle, every time the procedure starts. The processing of items following the structure  $AABB$  can easily be accomplished by counting the  $A$ 's and matching the number with the amount of  $B$ 's. Hence, if the positions in a sequence ( $S \rightarrow A_1B_1, S \rightarrow A_2A_1B_1B_2$ , etc.) are not explicitly defined as a dependency relation, the processing of this structure can be accomplished via counting.

In order to prevent such counting strategies each rule type used in the present study formulated a further requirement. The adjacent dependency rule included the requirement that each pair of adjacent  $A, B$  required a match with respect to the phonetic features [voice] and [place of articulation]. Similarly, the hierarchical dependency rule included the requirement that elements that represent one  $S$ , i.e.  $AB$  have to match with respect to [voice] and [place of articulation]. We will henceforth indicate this relation by the insertion of an index parameter ( $A_xB_x$ ). This additional rule describing the relationship between elements in a sequence, ensures that in the hierarchical dependency condition sequences are processed by participants in an embedded manner.

The processing of hierarchical dependency structures ( $A_2A_1B_1B_2$ ) and adjacent dependency structures ( $A_1B_1A_2B_2$ ) to our knowledge has not yet compared directly. The present study uses a within-subject design making such a direct comparison possible. The main question was, whether Broca's area would be engaged by the processing of hierarchical structures with rule guided dependencies between corresponding categories as compared to structures not comprising hierarchical dependencies. If so, we would be able to corroborate that the activation in Broca's area is indeed due to the processing of hierarchical dependencies, and not to a more simple processing strategy. A direct comparison of the two different grammar types for the correct sequences,

moreover, will allow to determine whether the expected activation in Broca's area is independent from any aspect of processing ungrammaticality which could be strategic in nature.

## Materials and methods

### Participants

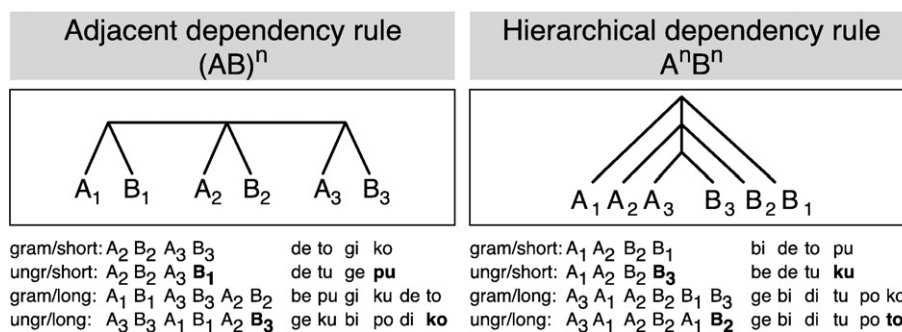
Sixteen native German speaking subjects participated in this study (9 male, mean age 25.5 years, SD = 3.7 years). They were right-handed and had a normal or corrected to normal vision. No participant had a known history of neurological, major medical, or psychiatric disorder. Prior to scanning, participants were informed about the potential risks and gave a declaration of consent. Due to technical problems, two of the 16 participants had to be excluded from the analysis.

### Stimuli

Sequences of consonant-vowel syllables were visually presented with one syllable at the time. Two types of categorizations of the syllables were applied. First, the syllables were assigned to two classes ( $A$  and  $B$ ), which were coded by different vowels. Class  $A$  syllables ended with /e/ or /i/ [be, bi, de, di, ge, gi] and class  $B$  syllables ended with /o/ or /u/ [po, pu, to, tu, ko, ku]. The two types of vowels (/e/ or /i/ within the  $A$ -category and /o/ or /u/ within the  $B$ -category) could occur randomly. Second, dependencies between pairs of  $A$ 's and  $B$ 's were coded by the phonetic features [voice] and [place of articulation]:  $A_1 - B_1 = b - p$ ;  $A_2 - B_2 = d - t$ ;  $A_3 - B_3 = g - k$  (see Fig. 1). The same syllables were used for both types of rules. The probability of occurrence of the syllable frequencies was balanced across positions in a sequence in order to prevent pattern learning. If one syllable occurred by chance more frequently in a certain position of one sequence, the participant could assume a rule behind this chunk. Hence, all syllables appeared with equal frequency in the experiment.

Using this stimulus material, a  $2 \times 2 \times 2$  factorial design was applied with the within-subject factors RULE TYPE (adjacent dependency rule, hierarchical dependency rule), LENGTH (short sequences, long sequences) and GRAMMATICALITY (grammatical sequences, ungrammatical sequences).

In the hierarchical rule, the structure of the syllable sequence followed the rule  $A^nB^n$ , the adjacent dependency rule was generated via the formula  $(AB)^n$  with the relation between dependent elements coded by phonetic features [voice] and [place of articulation]. For both types of rules the same syllables were selected, short (4 syllables) and long (6 syllables)



**Fig. 1.** General structure and examples of the two rule types. The adjacent dependency rule was generated by simple transitions between categories of consonant-vowel syllables. The hierarchical rule was produced by embeddings between the two syllable categories. Short and long sequences were applied. Violations of the structure were situated at the last 3 or 4 positions (short sequences) and at the last 4, 5, or 6 positions (long sequences). In the given example, the violations are placed at the fourths position for short sequences and at the sixth position for long sequences (bold letters).

sequences were applied, and grammatical and ungrammatical sequences were used (50% each).

#### *Hierarchical dependency rule*

An example for an embedded structure is the following:  $A_3A_2A_1B_1B_2B_3 = gi\ de\ be\ po\ tu\ ko$ . In order to prevent a mere pattern learning strategy, the positions of the dependent elements in a sequence were varied (six combinations for long sequences:  $A_3A_2A_1\ B_1B_2B_3$ ;  $A_2A_1A_3\ B_3B_1B_2$ ;  $A_3A_1A_2\ B_2B_1B_3$ ;  $A_2A_3A_1\ B_1B_3B_2$ ;  $A_1A_2A_3\ B_3B_2B_1$ ;  $A_1A_3A_2\ B_2B_3B_1$ ; and six combinations for short sequences:  $A_1A_2\ B_2B_1$ ;  $A_2A_1\ B_1B_2$ ;  $A_3A_2\ B_2B_3$ ;  $A_2A_3\ B_3B_2$ ;  $A_1A_3\ B_3B_1$ ;  $A_3A_1\ B_1B_3$ ). Furthermore,  $2^6 = 64$  combinations of the permutations of the categories  $A(e, i) \wedge B(o, u)$  were generated for long sequences and  $2^4 = 16$  category combinations for short sequences. Hence,  $64 \times 6 = 384$  items for long sequences and  $16 \times 6 = 96$  items for short sequences were computed in order to enforce the learning of the rules instead of learning syllable patterns. The testing of the correct processing of the embedded structure was implemented by introducing violations (ungrammatical sequences) into the structure. In order to avoid the likelihood of the participants focusing only on one particular position in the sequences, violations could occur at different positions of the sequences. Thus, violations occurred at one of the last two positions in short sequences (**AABB**), and at one of the last three positions in long sequences (**AAABBB**).

#### *Adjacent dependency rule*

An example for an adjacent dependency structure is the following:  $A_3B_3A_2B_2A_1B_1 = ge\ ko\ de\ to\ be\ po$ . As in the hierarchical rule, the order of sub-categories was changed systematically:  $A_2B_2A_1B_1A_3B_3$ ;  $A_1B_1A_3B_3A_2B_2$ ; etc. Likewise, for long sequences six variations of the positions and 64 combinations of permutations of the categories resulted in 384 possible syllable sequences. For short sequences, 96 possible syllable sequences were computed. Thus, in the adjacent dependency and in the hierarchical dependency rule the same amount of items were generated. Just as in the previous rule type, violations could occur at one of the last two positions in short sequences (**ABAB**), and one of the last three positions in the long sequences (**ABABAB**).

#### *Ungrammatical sequences*

In order to have the participants process both types of categorization, two classes of violations were employed for both rules. In case of *replacement violations*, class A and B syllables were replaced (e.g. adjacent dependency rule, short sequence:  $A_1B_1A_2A_2$ ; hierarchical dependency rule, long sequence:  $A_1A_2A_3B_3A_2B_1$ ). In case of *concatenation violations*, the plosive concatenations were violated (e.g. adjacent dependency rule, long sequence:  $A_1B_1A_2B_2A_3B_2$ ; hierarchical dependency rule, short sequence:  $A_1A_2B_2B_3$ ). Violations were generated as follows: In short syllable sequences, each of the two error types could occur at one of the last two positions, resulting in four possible combinations of the violation. All violations occurred in each of the 96 items. Hence,  $4 \times 96 = 384$  incorrect items were generated. In long sequences, the two error types could occur at one of the last three sequence positions. Accordingly,  $6 \times 384 = 2304$  incorrect items were generated for long sequences. For the learning and for the testing procedure different syllable sequences were randomly chosen from this item pool for each participant.

#### *Procedure*

Two experimental sets were performed. One set comprised the *learning* and *testing* of the adjacent dependency rule, the

other that of the hierarchical dependency rule. Hence, each participant learned and processed both of the rule types.

The learning of the rules took place 2 days prior to the scanning. The sets were separated by 7 days, and the order of the sets was balanced across participants. Each experimental set was subdivided into a *learning* and a *testing* session.

#### *Learning*

The learning period was subdivided into four parts which consisted of several blocks with different sequence lengths. First, sequences with two syllables were learned. Participants were instructed to extract the rule underlying the syllable sequences. At the beginning of each block, 10 grammatical sequences were presented. Afterwards, 5 grammatical and 5 ungrammatical sequences were shown in random order to which participants were required to respond at the end of each sequence with a button press indicating whether the sequences were grammatical or ungrammatical. Feedback was given for each sequence (i.e., “correct” or “incorrect” was presented on the screen). The first part of the training ended when participants correctly answered 90% of the trials in two successive blocks. Thereafter, sequences with four syllables were trained by the same procedure as with two items. Again, after participants reached a level of 90% of correct answers in two successive blocks, the third level of learning commenced. In the third part of learning, participants were inquired to learn sequences with six syllables until 90% of the items were correctly answered in two successive blocks. In the last part of the learning, participants were asked to judge grammaticality in 60 randomly presented sequences with four and six syllables. On average, the learning procedure took about 35 minutes.

#### *Testing*

During testing of the rules, fMRI measurements were conducted. Participants were instructed to judge whether the sequences were rule-based or not. A MR compatible response box with two buttons was used to record participants' judgment. Response was given with the right index or middle finger. As in the training session, feedback was also given here. For the testing period, 144 new sequences (half short, half long) were presented; half of them (36 for each length) were ungrammatical. The 36 ungrammatical sequences consisted of 18 *replacement violations* and 18 *concatenation violations* for both long and short sequences. In the short sequences, half of the violations (nine sequences each violation type) were situated at the third position and half of them were included at the fourth position. In the 36 long ungrammatical sequences, one third (six sequences each error type) were included at the fourth position, one third at the fifth, and one third at the sixth position. Additional 36 null events (fixation cross) were included in random order.

The start of a sequence was indicated by a fixation cross (500ms). Syllables were presented separately for 800ms with an inter-stimulus interval of 200ms between the syllables. After syllable presentation, participants could deliver their judgment for 1000ms, followed by a feedback for 500ms. Afterwards a fixation cross was shown for 1000ms. Trials started with a jitter of 0, 500, 1000, or 1500ms.

Response key assignment, and order of the grammar type (adjacent dependency rule first vs. hierarchical dependency rule first) were counterbalanced across participants.

#### *Steps taken to ensure hierarchical rule acquisition*

A couple of steps were taken to prevent alternative strategies like rote learning. First, the order of the category positions in a

syllable sequence was systematically changed, such that legal stimuli could have one of six different orders ( $A_3A_2A_1$   $B_1B_2B_3$ ;  $A_2A_1A_3$   $B_3B_1B_2$ ; etc.). Second, categories could comprise of an /e/ or /i/ (category A) or of an /o/ or /u/ (category B) resulting in  $2^6 = 64$  combinations of syllable sequences. All in all,  $64$  (category combinations)  $\times 6$  (possible positions) = 384 items were generated for long sequences and 96 items for short sequences. For testing 144 items were randomly chosen from that item pool. Finally, learning items and testing items were different (see also Stimuli section). As learning took on average about 8.5 blocks, with 10 grammatical items each, rote learning of every single syllable sequence could have not been accomplished within this short amount of time.

Counting of A-category items and matching it with B-category items were prevented by including two types of violations (see Procedure Testing), i.e. *replacement violation* (e.g.  $A_1A_2$   $B_1A_3$ ), and *concatenation violation* (e.g.  $A_1A_2$   $B_1B_3$ ). The error in the former violation type could be detected while counting As and matching the Bs (i.e. the last element is not a member of the B-category). However, counting would not be sufficient for error detection in the latter violation type, since the number of As equals the number of Bs. Real category matching and long distance processing must be applied in order to correctly reject this sequence. The two violation types were presented in random order and errors were always included at the end of a sequence. Thus, participants could not know from the beginning of a syllable sequence at which sequence they could apply a counting strategy. They always needed to use the category matching strategy in order to reach the criteria of 90% correct answers in two successive learning blocks.

Grammatical, ungrammatical, short, and long sequences were presented in random order. Syllables were presented separately and successively in the middle of the screen. Thus each syllable has to be kept in working memory to be matched with the corresponding category at the end of a sequence. Given that processing can be terminated earlier in violation conditions, lower processing costs for ungrammatical sequences was predicted. The lower processing costs should be reflected by less brain activation for the ungrammatical conditions compared to the grammatical conditions.

#### fMRI data acquisition

Imaging was performed at 3T Trio scanner (Siemens, Erlangen, Germany) equipped with the standard birdcage head coil. Stabilization cushions were used in order to reduce head motion. The anatomical and functional MRI protocol was identical in both sessions. For registration purpose, two sets of two-dimensional anatomical images were acquired for each participant immediately prior to the functional imaging. An MDEFT and an EPI-T1 sequence was used. T1-weighted MDEFT (Ugurbil et al., 1993) images (data matrix  $256 \times 256$ , TR = 1.3 seconds, TE = 7.4ms) were obtained with a non slice-selective inversion pulse followed by a single excitation of each slice (Norris, 2000). In addition, a set of T1-weighted spin-echo EPI images (TE 14ms, TR 3000ms) was taken with the same geometrical parameters (slices, resolution) and the same bandwidth was used for the fMRI data. A slice-selective inversion pulse was applied with an inversion time of 1200ms. Anatomical images were positioned parallel to AC–PC. Functional data were acquired from 16 axial slices (thickness = 3mm; gap = 0.6mm) using a gradient-echo EPI sequence with a TE = 30ms, flip angle =  $90^\circ$ , TR = 2000ms, acquisition bandwidth = 100kHz. The matrix acquired was  $128 \times 128$  with a FOV of 25.6cm, resulting in an in-plane resolution of  $2\text{mm} \times 2\text{mm}$ . One functional run with 780 volumes was measured.

#### Functional imaging data analysis

The fMRI data processing was performed using the software package LIPSIA (Lohmann et al., 2001). Functional data were motion-corrected offline with the Siemens motion correction protocol (Siemens, Erlangen, Germany). To correct for the temporal offset between the slices acquired in one scan, a cubic spline interpolation was applied. A temporal high-pass filter with a cutoff frequency of 1/30Hz was used for baseline correction of the signal and a spatial Gaussian filter with 5.65mm FWHM was applied. To align the functional data slices with a 3D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (3 rotational, 3 translational) was performed. The rotational and translational parameters were acquired on the basis of the MDEFT and EPI-T1 slices to achieve an optimal match between these slices and the individual 3D reference data set. The MDEFT volume data set was standardized to the Talairach stereotactic space (Talairach and Tournoux, 1988). The rotational and translational parameters were subsequently transformed by linear scaling to a standard size. The resulting parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system. The linear normalization was further improved by an additional nonlinear normalization.

The statistical evaluation was based on a least-squares estimation using the general linear model dealing with serially autocorrelated observations using pre-whitening of each voxel's time series (Worsley et al., 2002). The design matrix was generated with a box-car function, convolved with the hemodynamic response function. The model equation, including the observation data, the design matrix and the error term, was convolved with a Gaussian kernel of dispersion of 4 seconds FWHM to deal with the temporal autocorrelation. In the following, contrast images, i.e. estimates of the raw-score differences between the conditions, were generated for each participant. The single-participant contrast images were then entered into a second-level random effects analysis for each of the contrasts. The group analysis consisted of a one-sample *t*-test across the contrast images of all participants that indicated whether observed differences between conditions were significantly distinct from zero. Subsequently, *t*-values were transformed into Z-scores. To protect against false positive activations, only regions with Z-score greater than 3.09 ( $p < 0.001$ ) and with a volume greater than  $240\text{mm}^3$  (30 voxels) were considered (Forman et al., 1995). Figures show the resulting Z-map overlaid onto an anatomical reference image. The potential confound of a session effect (global BOLD signal differences across sessions) was taken into account within the general linear model. The design matrix consisted of the five experimental conditions (short/grammatical, short/ungrammatical, long/grammatical, long/ungrammatical, and null-event) for each of the two sessions (hierarchical dependency rule, adjacent dependency rule). The last two columns modeled the average activity in each session. Testing the session effect (via a *t*-test of the columns session one and session two) revealed no significantly activated brain region ( $p < .001$ , uncorrected). This result demonstrated that there was no significant difference in the global signal intensity.

#### Time course analysis

The time course analysis was performed on the basis of the results of the RULE TYPE contrast (hierarchical dependency rule, adjacent dependency rule), that was calculated in the second-level analysis. The results of this contrast revealed activity in the

left IFG and the adjacent rim of the ventral premotor cortex (BA 44/6) and in the right ventral premotor cortex (BA 6). The time course analysis was performed in the voxels that yielded the highest activation in the left BA 44/6 and the right BA 6. The shape and size of the two activated areas that survived a threshold of  $Z > 3.09$  were slightly different. The activated area in the left BA 44/6 comprised of 36 and in the right BA 6 of 43 adjacent voxels. The peak of activation in the left BA 44/6 was situated on the inferior precentral sulcus. In contrast, the activation maxima of the right-side region were located in the middle of the ventral premotor cortex. The time course analysis was employed to explore the influence of the factors LENGTH (short sequences, long sequences) and GRAMMATICALITY (grammatical sequences, ungrammatical sequences) on the hemodynamic response in Broca's area. Additionally, to detect a possible functional influence of the hemisphere, the right brain area corresponding to the peak coordinates of the right ventral premotor cortex was also included in the analysis. Hence, time course analysis was accomplished in the left inferior frontal gyrus (-46, 5, 16; BA 44/6) and the right ventral premotor cortex (46, -2, 16; BA 6). The onset time of all trials related to the specific variables was aligned in each participant. The time course of the individual preprocessed data was then averaged across participants. In the present study the activation peak slightly differed between long and short sequences. In order to take this latency effect into account, time courses were analyzed at four different time points. Time points 5, 7, 9, and 11 seconds were chosen, since the hemodynamic response is known for a temporal delay (approximately five seconds delay to peak) and an acquisition time of 2 seconds was used in the present study. Hence, starting at 5 seconds after stimulus onset, four time points every 2 seconds were selected (5, 7, 9, and 11). Paired *t*-tests were performed in both areas for each of the four factor combinations. For the significant effects, the partial effect size measure  $\omega^2$  was calculated.<sup>1</sup> The partial effect size reflects the amount of variance of the dependent variable that is declared by the independent variable (e.g., Hays, 1973) and ranges between 0 and 1. Note, that time steps were interdependent due to the temporal autocorrelation in the BOLD signal. Hence, an additional more conservative analysis was conducted in which calculations were performed at the peak of activation (Slotnick, 2005). Since the length of the stimulus presentation was unequal, after visual inspection the peak of the BOLD response was defined at 6 seconds after stimulus onset for short sequences and at 8 seconds after onset for long events.

**Results**

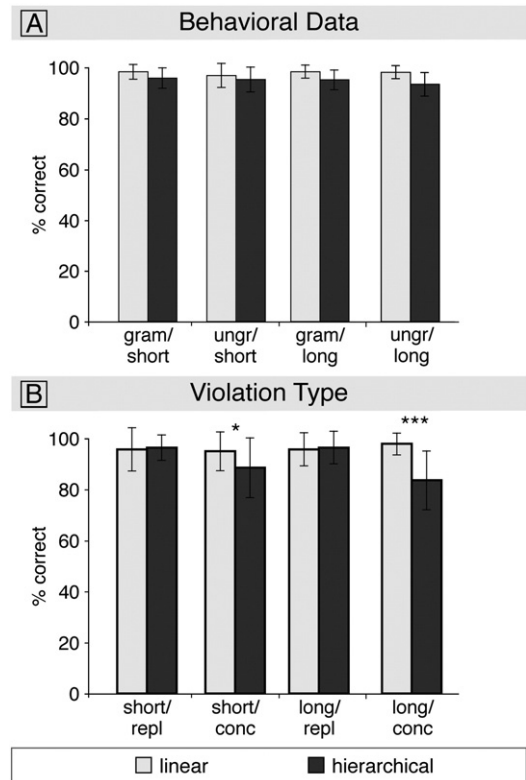
*Behavioral results*

In order to determine whether the learning of one of the two rule types takes longer than the other, a paired *t*-test was conducted at which the amount of learning blocks that were needed to reach the criteria was analyzed. There was no significant difference associated with the learning time [ $t(13) = 1.77$ , n.s.]. On average, participants needed 8.5 learning blocks (SD = 1.7) to reach the criteria of 90% correct answered sequences in two successive blocks. To investigate whether the order of learning of the two different rules had an influence on the performance of the participants, a separate analysis was accomplished. An ANOVA with the factors SESSION (first learning session, second learning

session) and RULE TYPE (hierarchical dependency rule, adjacent dependency rule) was carried out. This analysis revealed that the rule type was not a predictor of learning time in either the first or the second session. The knowledge of one rule did not influence the learning time of the other rule, i.e. no transfer effect was found [ $F(1,12) = 1.482$ , n.s.]. Additionally, a paired sample *t*-test was performed, comparing the learning time of the first session with the second session. The session effect was not significant [ $t(13) = 0.31$ , n.s.]; the number of learning blocks in the first session was found to be not significantly different from that needed in the second session.

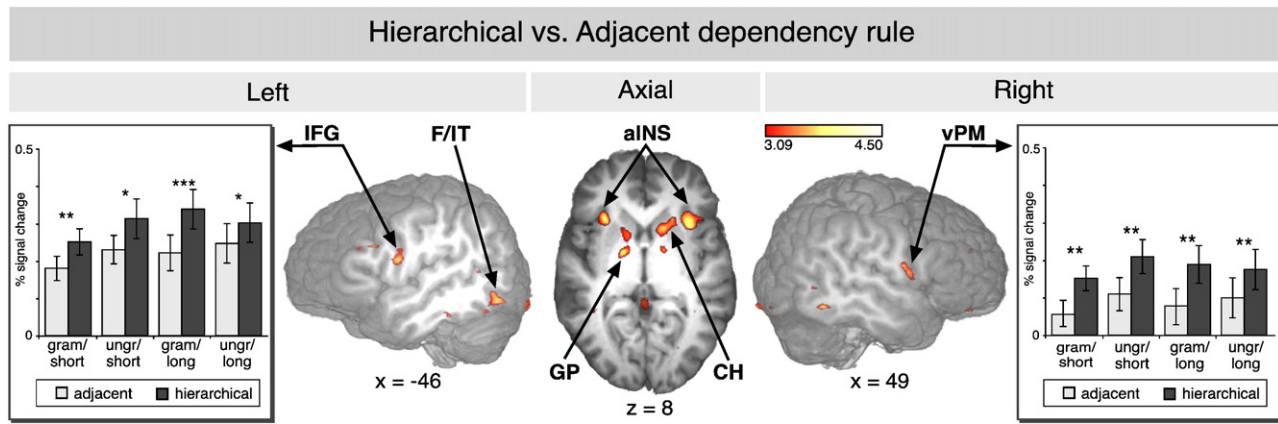
To investigate the performance in the *test* session, an ANOVA on the error rates was conducted, including the within-subject factors RULE TYPE (hierarchical dependency rule, adjacent dependency rule), LENGTH (short sequences, long sequences), and GRAMMATICALITY (grammatical sequences, ungrammatical sequences). Due to the delayed response mode used in the present study reaction times could not be analyzed. A significant main effect of RULE TYPE was found [ $F(1,13) = 12.97$ ,  $p < .01$ ], indicating that the hierarchical dependency rule induced slightly more errors (5.14%, SD = 3.4) than the adjacent dependency rule (2.16%, SD = 2.5; see Fig. 2A).

In order to ensure that syllable sequences were processed in a hierarchical way, the two types of structure violations, i.e. replacement and concatenation violations were analyzed. As described in the Procedure section, either a B-category was



**Fig. 2.** Behavioral data. A: Accuracy of hierarchical (dark gray) and adjacent dependency (light gray) rule (gram/short = grammatical, short items; ungr/short = ungrammatical, short; gram/long = grammatical, long; ungr/long = ungrammatical, long sequences). Note, that only the main effect of rule type was significant, no interaction effect occurred. The four combinations of the factors rule type are placed for illustrative reasons. B: Percentage of correct answered items for the two violation types in hierarchical (dark gray) and adjacent dependency (light gray) rule (short/repl = short sequences with replacement violation; short/conc = short sequences with concatenation violation; long/repl = long, replacement; long/conc = long concatenation error).  $N = 14$ , \* $p < .05$ , \*\*\* $p < .001$ .

<sup>1</sup> The calculation of the effect size was based on the formula  $\omega^2 = \frac{(t-1)}{(t-1)+n_{obs}}$ .



**Fig. 3.** BOLD response on the main effect hierarchical vs. adjacent dependency rule. Brain activation pattern elicited by the contrast vector of the hierarchical dependency rule versus the adjacent dependency rule. IFG = inferior frontal gyrus; F/IT = fusiform/inferior temporal gyrus; aINS = anterior insula; GP = globus pallidus; CH = caudate head; vPM = ventral premotor cortex. Left diagram: ROI analysis of the different variables obtained in the left BA 44/6. Right diagram: ROI analysis in the right BA 6. Hierarchical rule (dark gray) shows a higher BOLD response in comparison to adjacent dependency rule (light gray) in L BA 44/6 and R BA 6, in grammatical, short sequences (gram/short); ungrammatical, short (ungr/short); grammatical, long (gram/long); and ungrammatical, long sequences (ungr/long). Left-hemispheric ROIs revealed a higher hemodynamic response in comparison to right-hemispheric ROIs.  $N = 14$ ,  $*p < .05$ ,  $**p < .01$ ,  $***p < .001$ .

replaced by an A-category (replacement violation, e.g.  $A_1A_2B_2A_3$ ), or the concatenation between categories was erroneous (concatenation violation, e.g.  $A_1A_2B_2B_3$ ). In the latter case, two As were correctly followed by two Bs, but the  $B_3$  category did not correspond to the  $A_1$  category. If participants would use a simple counting mechanism, the concatenation violation should not be detected during processing of the sequence. Following this rationale, the detection of concatenation violations should be on chance level. To test this hypothesis, an ANOVA was conducted with the factors RULE TYPE (hierarchical dependency rule, adjacent dependency rule), ERROR TYPE (replacement violation, concatenation violation), and LENGTH (long sequences, short sequences). A marginally significant three-way interaction was found [ $F(1,13) = 3.70$ ,  $p = .07$ ]. The step down analyses revealed that this interaction was caused by the higher error rate of concatenation violations, only in the hierarchical dependency rule, which was more prominent for long sequences [ $t(13) = 4.62$ ,  $p < .001$ ], in comparison to short sequences [ $t(13) = 2.22$ ,  $p < .05$ ] (see Fig. 2B). This reflects that error rate for concatenation violations was not at chance level in the hierarchical dependency rule (13.5% errors). Hence, as participants linked the categories successfully, a simple counting mechanism can be excluded.

#### fMRI results: activation patterns

In the present study, a  $2 \times 2 \times 2$  design was applied with the factors RULE TYPE (hierarchical dependency rule, adjacent dependency rule), GRAMMATICALITY (grammatical sequences, ungrammatical sequences), and LENGTH (short sequences, long sequences). In order to investigate the three main effects, contrast images for RULE TYPE, GRAMMATICALITY, and LENGTH were computed.

The processing of hierarchical as compared to adjacent dependency rules (main effect of RULE TYPE) yielded activity in Broca's area and the adjacent rim of the ventral premotor cortex (see Fig. 3 and Table 1A).

We used the cytoarchitectonic probability map by Amunts et al. (1999) to further evaluate the location of this activated area (<http://www.bic.mni.mcgill.ca/cytoarchitectonics/>). Amunts et al. (1999) analyzed the microanatomical structure of 10 post-mortem brains, using an observer-independent algorithm for identifying borders between brain regions. The results of this analysis were spatially normalized and mapped to an individual

MNI reference brain. We compared the peak of activation (Tal:  $-46\ 5\ 16$ ; MNI:  $-46\ 4\ 18$ ) that was found to be activated in the main effect RULE TYPE with the cytoarchitectonic probability map. The result revealed a probability of 20% as being located within BA 44. This relatively low probability was observed because the peak of activation was situated at the posterior border of the pars opercularis. Microanatomically, this is a transition area between BA 44 and BA 6. However, most voxels of the activated area were located at the middle portion of pars opercularis (the peak was shifted to the posterior part of this activated area). The activated area ranged from  $-45$  to  $-52$  in the  $x$ -direction, from 1 to 16 on the  $y$ -axis, and from 4 to 19 on the  $z$ -axis. To conclude, the cytoarchitectonic probability map revealed that the area that we found to be activated comprised

**Table 1**

Anatomical areas, approximate Brodmann's Area (BA), mean Talairach coordinates ( $x$ ,  $y$ ,  $z$ ), and maximal  $Z$  values of the significant activations

	Brain region	BA	$x$	$y$	$z$	$Z_{\max}$	
(A) Main effect of RULE TYPE	<i>Hierarchical vs. adjacent dependency</i>						
	L inferior frontal G/precentral S	44/6	-46	5	16	3.72	
	R ventral premotor C	6	46	-2	16	3.70	
	L inferior frontal S	45/9	-34	28	22	3.80	
	L anterior insula		-28	18	4	4.23	
	R anterior insula		32	14	4	4.54	
	L fusiform/inferior temporal G	37	-42	-56	-18	4.20	
	R fusiform/inferior temporal G	19	30	-82	-18	3.64	
	L lingual G	17	-22	-98	-18	4.07	
	R lingual G	17	22	-98	-14	3.82	
	R caudate head		14	12	6	4.08	
	L globus pallidus lateralis		-16	-6	4	4.38	
	(B) Main effect of GRAMMATICALITY	<i>Ungrammatical vs. grammatical</i>					
		L anterior insula		-32	18	0	5.83
R anterior insula			32	20	4	5.78	
L superior temporal S		22/42	-42	-46	14	5.31	
R inferior occipital G		18	38	-84	-8	5.06	
L fusiform G		37	-38	-50	-14	4.80	
L thalamus (posterior lateral Ncl)			-8	-22	16	5.00	
R thalamus (ventral lateral Ncl)			12	-10	12	4.33	
R thalamus (pulvinar Ncl)			14	-30	12	4.60	
R cerebellum			12	-68	-20	4.88	
(C) Main effect of LENGTH	<i>Long vs. short</i>						
	L anterior insula		-26	18	10	4.23	
	R anterior insula		28	18	4	5.09	

Main effects.

L, left hemisphere; R, right hemisphere; G, gyrus; S, sulcus; Ncl, nucleus.

of parts of the left inferior frontal gyrus together with parts of the adjacent rim of the ventral premotor cortex (BA 44/6). Additional activation was found in the right ventral premotor cortex, the bilateral anterior insula, the left inferior frontal sulcus, the bilateral fusiform and lingual gyrus, and bilateral sub-cortical activations (including caudate nucleus).

Processing ungrammatical as compared to grammatical sequences (main effect of GRAMMATICALITY) elicited several bilateral activations (see Table 1B), including the anterior insula, the left superior temporal sulcus, the right inferior occipital gyrus, the right cerebellum, and the bilateral thalamus.

Finally, processing long versus short sequences (main effect of LENGTH) was correlated with bilateral activity in the anterior insula (see Table 1C).

In order to clarify a putative interaction between rule type and sequence violation, additional contrasts between hierarchical versus adjacent dependency rules were conducted for the grammatical sequences. As a result, activity in Broca's area and the right ventral premotor cortex remained in this comparison. In contrast, Broca's area was not engaged when conducting the comparison of hierarchical versus adjacent dependencies for the ungrammatical sequences (see Table 2A). This finding indicates that the activity in the left BA 44 and right BA 6 is due to the processing of grammatical sequences.

An additional analysis of the fMRI data was conducted to rule out the activation in Broca's area was confounded with task difficulty. Behavioral data indicated that participants conducted slightly more errors in the hierarchically structured trials (5% errors) in comparison to the adjacent dependency rule (2% errors). This difference in performance given to the low absolute values of this difference, it seemed unlikely that it had a significant impact on the BOLD response. In order to rule out the possible confound of task difficulty as the main cause for the activity in Broca's area we conducted an analysis which included the individual performance values for each condition as an additional parameter in the general linear model. If this parameter would declare the variance of the BOLD response in Broca's area, than the performance difference rather than rule type (adjacent dependency versus hierarchical dependency) would be the cause of activity in this area. However, no brain

region was found to be activated when testing this parameter. Accordingly, difficulty can be excluded as a possible cause for the activations in the brain regions found in this study.

*fMRI results: time courses*

The main result, namely the difference between hierarchical and adjacent dependencies was further evaluated in a time course analysis.

The RULE TYPE contrast yielded activity in predicted region of interest (ROI), i.e. the left BA 44/6 and also the right BA 6. In order to specify the possible influence of the factors GRAMMATICALITY and LENGTH on the activity detected for hierarchical sequences in contrast to sequences following the adjacent dependency rule, time courses were analyzed in the left BA 44/6. Since the right ventral premotor cortex was also identified, this region was included in the analysis. The paired *t*-test between left BA 44/6 and right ventral premotor cortex illustrated a general higher BOLD response for the left-sided activation in comparison to the right-hemisphere activity [ $t(13) = 2.41, p < 0.05$ ].

The higher hemodynamic response to hierarchical sequences in comparison to adjacent dependency sequences was shown for grammatical, ungrammatical, short, and long sequences in both hemispheres: short/grammatical sequences (left: [ $t(13) = 4.06, p < .01, \omega^2 = .1$ ], right: [ $t(13) = 4.53, p < .01, \omega^2 = .11$ ]), short/ungrammatical (left: [ $t(13) = 2.53, p < .05, \omega^2 = .05$ ], right: [ $t(13) = 3.94, p < .01, \omega^2 = .09$ ]), long/grammatical (left: [ $t(13) = 5.32, p < .001, \omega^2 = .13$ ], right: [ $t(13) = 4.50, p < .01, \omega^2 = .11$ ]), and long/ungrammatical sequences (left: [ $t(13) = 2.63, p < .05, \omega^2 = .05$ ], right: [ $t(13) = 3.29, p < .01, \omega^2 = .07$ ]). As expected the effect size in Broca's area was larger for grammatical sequences (short:  $\omega^2 = .1$ ; long:  $\omega^2 = .13$ ) than for ungrammatical sequences (short and long:  $\omega^2 = .05$ ). This indicates a general higher hemodynamic response for hierarchical, grammatical sequences in comparison to hierarchical ungrammatical sequences (see Fig. 3). This pattern was also shown for the more conservative analysis at the peak of activation. In this analysis, paired *t*-tests were performed at time point six after stimulus onset for short sequences and time point eight for long sequences. Within each of the four combinations between the two factors GRAMMATICALITY and LENGTH, the simple main effects were significant. This was shown for the left-side ROI [ $3.04 \leq t(13) \leq 5.34, .001 \leq p \leq .01$ ] and for the right-side ROI [ $2.21 \leq t(13) \leq 5.12, .001 \leq p \leq .05$ ].

**Table 2**

Anatomical areas, approximate Brodmann's Area (BA), mean Talairach coordinates (*x, y, z*), and maximal *Z* values of the significant activations: effect of RULE TYPE: separate analysis for grammatical and ungrammatical sequences

(A) Grammatical: hierarchical vs. adjacent dependency					
L inferior frontal G	44/6	-46	4	16	3.84
R ventral premotor C	6	46	-2	16	3.81
L anterior Insula		-26	18	4	4.15
R anterior Insula		32	14	4	4.37
L inferior frontal S	45/9	-34	28	22	3.89
L fusiform G/inferior temporal G	37	-36	-56	-20	4.35
L caudate body		-18	-6	22	3.99
R cerebral peduncle		16	-18	-8	4.79
L lingual G	18	-22	-98	-18	4.40
R lingual G	18	22	-98	-14	4.03
L globus pallidus lateralis		-16	-6	2	4.26
(B) Ungrammatical: hierarchical vs. adjacent dependency					
L anterior insula		-28	18	4	4.24
R anterior insula		30	12	8	4.63
L middle frontal G	9	-34	28	22	3.68
L globus pallidus lateralis		-16	-6	4	4.37
R caudate head		14	12	6	4.16
L fusiform/inferior temporal G	19/37	-50	-68	-12	4.09
L lingual G	18	-22	-98	-18	3.56

L, left hemisphere; R, right hemisphere; G, gyrus; S, sulcus; Ncl, nucleus.

**Discussion**

The present study investigated the brain basis of processing hierarchical structures by directly comparing the processing of hierarchically structured syllable sequences with sequences following an adjacent dependency rule. Due to the subtraction design, several cognitive processes that were engaged during the processing of both structures could be controlled for. The main goal of this study was to explore the brain basis of grammatical rule use, rather than the learning of grammatical rules. The processes underlying the application of the hierarchical rule in the study at hand may thus differ from those implicated in implicit or incidental learning paradigms.<sup>2</sup> Here

<sup>2</sup> Note also, that in the present study visual presentation was applied, in order to stay comparable to numerous other language processing and artificial grammar learning experiments (cf. Tettamanti et al., 2002; Musso et al., 2003; Fiebach et al., 2005). Even if one could assume common grounds between written and spoken grammar processing, our conclusions remain restricted to the former.

we focus on brain responses to the processing of two different rule types established by prior learning. Activation as a function of the different rules was found in Broca's area, but also some other areas which, however, partly seem to reflect aspects of grammaticality and general processing load (length). Note that from the theoretical linguistic perspective one might also argue that both artificial grammar rules at hand could be considered as context-sensitive grammars, since both require a match between two constituents (cf. Chomsky, 1959). Thus, both rule types might be on the same level of the Chomsky hierarchy. Accordingly, the observed activation pattern in the contrast of hierarchical versus adjacent dependency rule may not be taken to reflect grammatical complexity, but different dependency types involving different verbal working memory load.

In the following we will first discuss activation in Broca's area and then turn for the other activation observed in the present study.

### Broca's area

The processing of the hierarchical dependency rule evoked higher BOLD response in Broca's area extending into the left premotor cortex compared to the adjacent dependency rule. An activation increase was also shown in the right ventral premotor cortex. The hemodynamic response in left BA 44/6 was higher than in the right BA 6, suggesting a left-dominant activation pattern. It is important to note that BA 44/6 remains significantly activated even when excluding ungrammatical items.

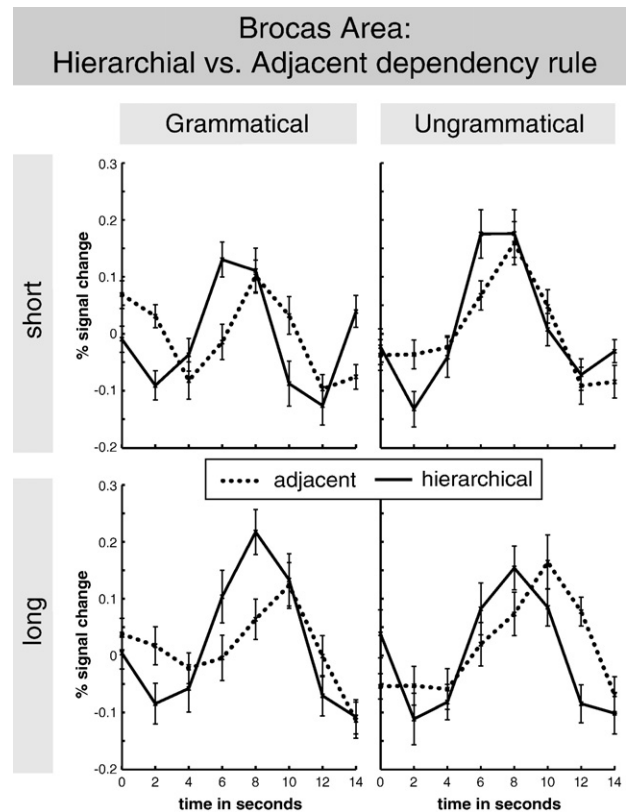
This finding indicates that Broca's area is particularly engaged in processing hierarchical as compared to non-hierarchical grammars, a finding further supporting the idea formulated in a previous study (Friederici et al., 2006a). While this notion was already suggested by the earlier study, the present experiment used a more elaborated grammar to test Broca's role in the processing of hierarchical structures more rigorously. In the present study we specifically aimed to exclude the possible confound of a simple counting mechanism (cf. Perruchet and Rey, 2005; de Vries et al., 2007).

To this end, the grammar of the present study was designed to build up defined long distance dependencies between respective elements of the categories A and B. The embedded nature of these sequences at hand is reflected by the dependency of categories over non-adjacent elements (e.g.,  $A_3A_2A_1B_1B_2B_3$ ). In order to validate an embedded type of processing instead of a simple counting strategy, errors that violated the matching of categories were included. Analysis of the behavioral data revealed an increased amount of errors conducted for concatenation violations (13.5%) in comparison to replacement violation (3.5%) (see Fig. 2B) suggesting a higher processing difficulty of the former error type. Importantly, however, even though concatenation violations caused more errors than replacement violations grammaticality judgment was not at chance level as predicted by a counting strategy. Still, one cannot exclude that for some items a counting or other strategies could be applied. For processing of items with replacement violations (e.g.,  $A_1A_2A_3B_1$ ) one could count the number of items per category and once the number is not the same for each category, the participant could have judged those sequences correctly as ungrammatical. However, this only holds for the replacement violation (25% of all sequences). In order to successfully process the other 75% of items one needs to access the concatenation strategy. Critically, we showed that the activity in Broca's area remains significant even when all ungrammatical items, comprising of concatenation and

replacement violations, were excluded. Hence, the activity in Broca's area could not be explained by simple error detection strategies, but by hierarchical processing.

Broca's area and the right ventral premotor cortex were significantly more activated for the processing of the hierarchical rule vs. the adjacent rule even when excluding ungrammatical sequences. Exclusion of grammatical sequences, by contrast, did not reveal any significant activity in these areas for the comparison of the hierarchical vs. the non-hierarchical condition. The time course analysis showed a significant difference in the BOLD response between the hierarchical and the adjacent dependency rule for grammatical and for ungrammatical items (see Fig. 4). However, the effect size for ungrammatical sequences was much smaller than for grammatical sequences. This result revealed that it is not violation detection per se that caused activity in Broca's area, but that the activity rather correlated with intrinsic differences between the two grammar types, namely hierarchical versus non-hierarchical.

From a broader perspective, the present findings on left BA 44/6 are in line with the literature on language-like artificial grammar processing (Opitz and Friederici, 2003, 2004; Friederici et al., 2006a) indicating this region's involvement in the processing of complex syntactic structures. Moreover, a number of fMRI studies investigating syntax processing in natural language also reported activity in Broca's area (Stromswold et al., 1996; Caplan, 2001; Moro et al., 2001; Röder et al., 2002; Ben-Shachar et al., 2003; Musso et al., 2003; Fiebach et al., 2005; Bahlmann et al., 2007). In natural language processing, the



**Fig. 4.** Time course analysis in BA 44/6. Averaged signal change of the hemodynamic response associated with each of the eight conditions (error bars correspond to standard errors across participants). Solid line: hierarchical dependency rule. Dotted line: adjacent dependency rule. The hierarchical dependency rule showed a higher BOLD response in all conditions. The difference of signal change between hierarchical and adjacent dependency rule is higher for grammatical (long and short) as compared to ungrammatical (long and short) conditions.



increase of activation in Broca's area after syntactic manipulations in sentences was interpreted as correlate of higher syntactic complexity (Caplan, 2001), higher syntactic working memory (Fiebach et al., 2005), or processing costs due to syntactic movement in hierarchical structures (Grodzinsky, 2000; Ben-Shachar et al., 2003; Santi and Grodzinsky, 2007), hierarchical structure building (Friederici, 2004; Friederici et al., 2006a), or action and language related hierarchical processing (Tettamanti and Weniger, 2006). Furthermore, procedural memory (Ullman, 2004, 2006), unification (Hagoort, 2005) and synthesizing (Ben Shalom and Poeppel, 2008), have been discussed as functional underpinnings.

In the present work, the observed activity in Broca's area was associated with the processing of a hierarchically structured sequences in which the dependent elements were linked by a hierarchical dependency relation. The processing of this structural dependency particularly engaged BA 44 and the adjacent rim of the ventral premotor cortex. Hence, in line with our previous study (Friederici et al., 2006a) we suggest that the activity in this regions was correlated with hierarchical structure building. Syntactic movement related dependencies in natural language processing also appear to engage Broca's area although a slightly more anterior portion, i.e. BA 45 (Santi and Grodzinsky, 2007). Whether the different neuroanatomical location of the activation within either BA 44 (present study) and BA 45 (Santi and Grodzinsky, 2007) for the processing of long distance dependency relations, is due to different language types (artificial vs. natural) or can be attributed to different dependency types (hierarchical vs. movement) will have to be clarified in further studies.

Note that the present results remain neutral as to whether Broca's area was engaged due to the fact that language-specific regions are exploited for the processing of language-like stimuli or, conversely, whether natural grammar must be conceived of as one of many possible hierarchical structures to be processed in Broca's area (e.g. Koehlin and Jubault, 2006). Rather, the notion put forward here is that Broca's area subserves the processing of hierarchical structures in the domain of grammar.

#### *Other activations*

There were also a number of other brain regions found to be activated for the different comparisons. Crucially, for the hierarchical vs. adjacent dependency comparisons in the grammatical sequences a number of additional brain areas were found. However, as the majority of these were also found to be significant in other comparisons these cannot be attributed to the aspect of hierarchy as such. These areas are, the insulae bilaterally which were effective in all comparisons (A–C, Table 1, A–B Table 2), the fusiform gyrus which showed an effect for all those comparisons including hierarchical versus adjacent dependency rule (A, B Table 1, B Table 2), and different nuclei of the basal ganglia which were effective in this latter comparisons.

#### *Insula*

In the bilateral dorsal anterior insula all three main effects, namely RULE TYPE, GRAMMATICALITY, and LENGTH were correlated with activity suggesting that this region is not specific for the processing of hierarchical structures. The anterior insula is connected with several cortical and sub-cortical brain areas (Preuss and Goldman-Rakic, 1989). Numerous cognitive and autonomous functions were associated with this region, including autonomic functions, subjective feeling, breathing, temporal

processing, cognitive control, or language processing (see Augustine, 1996 for an overview). With respect to language processing, an involvement of the anterior insula has been reported in particular for speech and auditory processing. For example, lesions in the anterior insula are correlated with impairments in the processing of speech (Kreisler et al., 2000; Saygin et al., 2004). Bamiou et al. (2003) discuss in their review the anterior insula as a region involved in the integration of auditory information with other associative functions. In a recent meta-analysis, Wager and Barrett (2004) suggested the anterior dorsal insula to particularly reflect the mental effort of a cognitive task which is accompanied by autonomic reactions. According to this more general notion, the insula activation in our tasks could be also due to an autonomous reaction triggered by a signal for optimizing ongoing task performance. Our behavioral data confirm this interpretation, as error rates were higher for hierarchical sequences compared to adjacent dependency sequences. In contrast, no behavioral differences were recorded for ungrammatical vs. grammatical and long vs. short sequences. Note, that both rule types were processed to level of high proficiency (overall only 3.65% conducted errors). The performance differences between grammatical vs. ungrammatical and long vs. short sequences could be diminished, due to the over-learned character of the stimuli. A particularly interesting question for further investigations would be, whether mental effort is the key to the insula activity in cognitive demanding tasks.

#### *Fusiform gyrus*

The left middle fusiform gyrus showed increased levels of activation in the main effect of RULE TYPE and in the main effect of GRAMMATICALITY. The peak of activation was situated in an area labeled as visual word form area. This region was consistently engaged during reading of words and readable pseudowords and is therefore taken to be involved in abstract letter identification due to the invariance of visual features (e.g.; Cohen et al., 2000). The specificity of that area has subsequently been questioned by several authors, as this region is also known to show an increased BOLD response to auditory word processing (Booth et al., 2002) and tactile (Braille reading) processing (Buchel et al., 1998). In the present study, this area was more activated for readable syllable sequences that followed a hierarchical dependency rule than for an adjacent dependency rule, and for ungrammatical syllable sequences than for grammatical sequences. The present findings could be reconciled with the above mentioned studies by assuming a recognition process based on pattern matching and expectation of an upcoming syllable form which would be more difficult for the hierarchical dependency rule than for adjacent dependency rule, and also more difficult for ungrammatical than for grammatical ones.

#### *Sub-cortical structures*

The basal ganglia were sensitive to different factors. The right caudate head and the left globus pallidus revealed higher hemodynamic response to the processing of the hierarchical compared to the adjacent dependency rule. Interestingly, when excluding ungrammatical sequences, increased activity was observed in the left caudate body and the left globus pallidus, whereas when excluding grammatical sequences activity was observed in the right caudate head and the left globus pallidus. These comparisons clearly indicate a differential function of the different substructures of the basal ganglia.

The involvement of the basal ganglia in language processing has long been discussed (Crosson, 1985). Their specific role,

however, remains unclear. While Ullman (2004) proposed that the basal ganglia together with Broca's area support procedural aspects of grammar, others assumed that the basal ganglia are rather involved in controlled grammatical processes (e.g. Friederici et al., 2003; Kotz et al., 2003). The present findings suggest that substructures of the left ganglia (left caudate body, left globus pallidus) are involved in on-line grammatical processes. The left globus pallidus, in particular, was found to be activated as a function of grammar type independent of its grammaticality which suggests its special role in grammatical processes.

Error detection, in contrast, seems to be subserved by *thalamic structures* bilaterally as these were found to be activated more for ungrammatical than for grammatical structures.

## Conclusion

The present study indicates that Broca's area supports the processing of hierarchical dependencies during artificial grammar processing. Broca's area subserves this function as a part of a neural network of natural language processing which involves several other brain areas.

## Acknowledgment

The authors would like to thank D. Yves von Cramon for providing his neuroanatomical expertise.

## References

- Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H.B.M., Zilles, K., 1999. Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 412 (2), 319–341.
- Augustine, J.R., 1996. Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Res. Rev.* 22, 229–244.
- Bahlmann, J., Rodriguez-Fornells, A., Rotte, M., Münte, T.F., 2007. An fMRI study of canonical and non-canonical word order in German. *Hum. Brain Mapp.* 28, 940–949.
- Bamiou, D.E., Musiek, F.E., Luxon, L.M., 2003. The insula (Island of Reil) and its role in auditory processing. Literature review. *Brain Res. Rev.* 42, 143–154.
- Ben-Shachar, M., Hendlar, T., Kahn, I., Ben-Bashat, D., Grodzinsky, Y., 2003. The neural reality of syntactic transformations: evidence from functional magnetic resonance imaging. *Psychol. Sci.* 14, 433–440.
- Ben Shalom, D., Poeppel, D., 2008. Functional Anatomic models of language: assembling the pieces. *Neuroscientist* 14, 119–127.
- Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T.D., Mesulam, M.M., 2002. Functional anatomy of intra- and cross-modal lexical tasks. *Neuroimage* 16, 7–22.
- Buchel, C., Price, C., Friston, K., 1998. A multimodal language region in the ventral visual pathway. *Nature* 394, 274–277.
- Caplan, D., 2001. Functional neuroimaging studies of syntactic processing. *J. Psycholinguist. Res.* 30, 297–320.
- Chomsky, N., 1959. On certain formal properties of grammars. *Inf. Control* 2, 137–167.
- Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M.A., Michel, F., 2000. The visual word form area – spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123, 291–307.
- Corballis, M.C., 2007. Recursion, language, and starlings. *Cogn. Sci.* 31, 697–704.
- Crosson, B., 1985. Subcortical functional in language: a working model. *Brain Lang.* 25, 257–292.
- de Vries, M.H., Monaghan, P., Knecht, S., Zwitserlood, P., 2007. Syntactic structure and artificial grammar learning: the learnability of embedded hierarchical structures. *Cognition* 107, 763–774.
- 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. *Hum. Brain Mapp.* 24, 79–91.
- Fitch, W.T., Hauser, M.D., 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.
- Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI) – use of a cluster-size threshold. *Magn. Reson. Med.* 33, 636–647.
- Friederici, A.D., 2004. Processing local transitions versus long-distance syntactic hierarchies. *Trends Cogn. Sci.* 8, 245–247.
- Friederici, A.D., Kotz, S.A., Werheid, K., Hein, G., von Cramon, D.Y., 2003. Syntactic comprehension in Parkinson's disease: investigating early automatic and late integrational processes using event-related brain potentials. *Neuropsychology* 17, 133–142.
- 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. *Proc. Natl. Acad. Sci. U. S. A.* 103, 2458–2463.
- Friederici, A.D., Fiebach, C.J., Schlesewsky, M., Bornkessel, I., von Cramon, D.Y., 2006b. Processing linguistic complexity and grammaticality in the left frontal cortex. *Cereb. Cortex* 16, 1709–1717.
- Gentner, T.Q., Fenn, K.M., Margoliash, D., Nusbaum, H.C., 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.
- Grodzinsky, Y., 2000. The neurology of syntax: language use without Broca's area. *Behav. Brain Sci.* 23, 1–21 discussion 21–71.
- Hagoort, P., 2005. On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9, 416–423.
- Hauser, M.D., Chomsky, N., Fitch, W.T., 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
- Hauser, M.D., Barner, D., O'Donnell, T., 2007. Evolutionary linguistics: a new look at an old landscape. *Lang. Learn. Dev.* 3, 101–132.
- Koechlin, E., Jubault, T., 2006. Broca's area and the hierarchical organization of human behavior. *Neuron* 50, 963–974.
- Hays, W.L., 1973. *Statistics for the Social Sciences*, 2nd ed. Holt, Rinehart and Winston, New York.
- Kotz, S.A., Frisch, S., von Cramon, D.Y., Friederici, A.D., 2003. Syntactic language processing: ERP lesion data on the role of the basal ganglia. *J. Int. Neuropsychol. Soc.* 9, 1053–1060.
- Kreisler, A., Godefroy, O., Delmaire, C., Debachy, B., Leclercq, M., Pruvot, J.P., Leys, D., 2000. The anatomy of aphasia revisited. *Neurology* 54, 1117–1123.
- Lohmann, G., Müller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., Zysset, S., von Cramon, D.Y., 2001. LIPSIA—a new software system for the evaluation of functional magnetic resonance images of the human brain. *Comput. Med. Imaging Graph.* 25, 449–457.
- Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S.F., Fazio, F., 2001. Syntax and the brain: disentangling grammar by selective anomalies. *Neuroimage* 13, 110–118.
- Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Buchel, C., Weiller, C., 2003. Broca's area and the language instinct. *Nat. Neurosci.* 6, 774–781.
- Norris, D.G., 2000. Reduced power multislice MDEFT imaging. *J. Magn. Reson. Imaging* 11, 445–451.
- Opitz, B., Friederici, A.D., 2003. Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *Neuroimage* 19, 1730–1737.
- Opitz, B., Friederici, A.D., 2004. Brain correlates of language learning: the neuronal dissociation of rule-based versus similarity-based learning. *J. Neurosci.* 24, 8436–8440.
- Perruchet, P., Rey, A., 2005. Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *PsycholBull. Rev.* 12, 307–313.
- Pinker, S., Jackendoff, R., 2005. The faculty of language: what's special about it? *Cognition* 95, 201–236.
- Preuss, T.M., Goldman-Rakic, P.S., 1989. Connections of the ventral granular frontal cortex of macaques with perisylvian premotor and somatosensory areas: anatomical evidence for somatic representation in primate frontal association cortex. *J. Comp. Neurol.* 282, 293–316.
- Röder, B., Stock, O., Neville, H., Bien, S., Röslér, 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.
- Santi, A., Grodzinsky, Y., 2007. Working memory and syntax interact in Broca's area. *NeuroImage* 37, 8–17.
- Saygin, A.P., Wilson, S.M., Dronkers, N.F., Bates, E., 2004. Action comprehension in aphasia: linguistic and non-linguistic deficits and their lesion correlates. *Neuropsychologia* 42, 1788–1804.
- Slotnick, S.D., 2005. Spatial working memory specific activity in dorsal prefrontal cortex? Disparate answers from fMRI beta-weight and timecourse analysis. *Cogn. Neuropsychol.* 22, 905–920.
- Stromswold, K., Caplan, D., Alpert, N., Rauch, S., 1996. Localization of syntactic comprehension by positron emission tomography. *Brain Lang.* 52, 452–473.
- Talairach, J., Tournoux, P., 1988. *Co-planar Stereotactic Atlas of the Human Brain*. Thieme, New York.
- Tettamanti, M., Weniger, D., 2006. Broca's area: a supramodal hierarchical processor? *Cortex* 42, 491–494.
- Tettamanti, M., Alkadhi, H., Moro, A., Perani, D., Kollias, S., Weniger, D., 2002. Neural correlates for the acquisition of natural language syntax. *Neuroimage* 17, 700–709.
- Ugurbil, K., Garwood, M., Ellermann, J., Hendrich, K., Hinke, R., Hu, X., Kim, S.G., Menon, R., Merkle, S., Ogawa, S., Salmi, R., 1993. Imaging at high magnetic fields: initial experiences at 4T. *Magn. Reson. Q.* 9, 259.
- Ullman, M.T., 2004. Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92, 231–270.
- Ullman, M.T., 2006. Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex* 42, 480–485.
- Wager, T.D., Barrett, L.F., 2004. From affect to control: functional specialization of the insula in motivation and regulation. Published online at [PsycExtra](#).
- Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., Evans, A.C., 2002. A general statistical analysis for fMRI data. *Neuroimage* 15, 1–15.