Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules

Bertram Opitz a, b, * and Angela D. Friederici b

a Experimental Neuropsychology Unit, Saarland University, Saarbrücken, Germany
b Max Planck Institute of Cognitive Neuroscience, Leipzig, Germany

Received 8 November 2002; revised 12 February 2003; accepted 20 March 2003

Abstract

One of the most influential views on the hippocampal function suggests that this brain region is critically involved in relational memory processing, that is, binding converging inputs to mediate the representation of relationships among the constituents of episodes. It has been proposed that this binding is automatic and obligatory during learning and remembering. In addition, neuroimaging studies have highlighted the importance of the prefrontal cortex in learning, memory, and language processing. However, the posited importance of hippocampal–prefrontal interaction remains to be empirically tested. In the present study we used functional magnetic resonance imaging to examine in detail this interaction by assessing learning-related changes in hemodynamic activity during artificial language acquisition. It has been shown previously that artificial grammar systems might be learned by evaluating pattern-based relations in word sequences and generalizing beyond specific word order, that is, rule abstraction. During scanning, participants learned an artificial language whose miniature grammar meets the universal principles of a natural language. Increased proficiency level of the artificial language is associated with decreased left hippocampal activity. In contrast, we observed an increased recruitment of the left inferior frontal gyrus (Broca’s area), a region that contributes to syntax processing in natural language. The present results, therefore, indicate a learning-related change in brain circuitry underlying relational processes of language learning, with a transition from a similarity-based learning system in the medial temporal lobes to a language-related processing system in the left prefrontal cortex.

© 2003 Elsevier Science (USA). All rights reserved.

Keywords: Learning; Hippocampus; Artificial grammar; fMRI

Introduction

Recent studies have begun to use functional neuroimaging techniques to describe the brain activity that occurs during human learning. One important view on learning, gaining considerable interest over the past years, proposes two distinct processes (Knowlton and Squire, 1996; Smith et al., 1998). According to this view, individuals either learn item-specific information on the basis of superficial similarity to some exemplar or learn by abstracting a representation of the underlying rules. Rule-based learning has been studied extensively by means of an artificial grammar-learning task. An artificial grammar system, first introduced by Reber (1967), reflects a set of arbitrary (nonlinguistic) rules governing the concatenation of letter strings. Usually subjects exposed to such grammar systems learn to categorize strings as grammatical (i.e., conforming to the rules) or ungrammatical and after some time of exposure perform this categorization task with an accuracy greater than chance (Reber, 1967). It has been argued, however, that learning mechanisms on the basis of superficial similarity may well explain the learning of such artificial grammars (Shanks, 1995). In one study it was shown that subject’s knowledge of parts of the grammar strings as assessed by a recognition procedure was sufficient to account for the observed performance on the artificial grammar-learning task (Perruchet and Pacteau, 1990). Despite such experiments that demonstrate that knowledge of exemplar-specific infor-
mation is sufficient to support some learning, this cannot explain the acquisition of abstract or rule-based knowledge in more complex artificial grammar paradigms. This was demonstrated by a nonlinguistic artificial grammar-learning experiment (Gomez and Schvaneveldt, 1994) in which subjects were exposed to one subset of strings generated by the artificial grammar during learning but to another subset generated by the same grammar during test. Despite the change in vocabulary subjects were able to discriminate grammatical from ungrammatical strings suggesting that subjects had extracted some aspect of the grammatical rules beyond specific exemplars. Thus learning of this type cannot solely be based on superficial similarities but necessarily requires abstraction abilities. Neural evidence for abstraction mechanisms has been provided by a number of studies employing functional neuroimaging techniques (Fletcher et al., 2001; Seger et al., 2000; Strange et al., 2001). For instance in the study by Strange et al. (2001) subjects made grammaticality judgments to exemplar letter strings for which the governing nonlinguistic grammar rule was periodically changed. Bilateral anterior prefrontal cortices were selectively engaged following abstract rule change but not following a change in exemplars conforming to the same rules. Interestingly, this latter change of superficial features modulated activation in the left hippocampus pointing to a differentiation between similarity-based learning and rule abstraction. In contrast, artificial grammar learning has been shown to be independent of the medial temporal lobe memory system in neuropsychological studies with patients suffering from global amnesia. Knowlton and Squire (1994, 1996) found that amnesic patients exhibit normal performance in an artificial grammar-learning task but were impaired on recognition judgments about the items used. Thus, what neural systems may subserve rule-based learning is still an open issue.

One aspect that has received little attention in previous studies is the potential difference between different artificial grammar systems. Despite apparent similarities between artificial grammars of the Reber type (Reber, 1967) and those used to study real language abilities in infants and adults (Saffran et al., 1996; Gomez and Gerken, 2000), recent electrophysiological and neuroimaging evidence emphasizes the vast differences between the two. Learning of nonlinguistic artificial grammars has been shown to activate a variety of cortical brain areas distinct from those areas activated by natural language processing. Although areas activated by artificial grammar learning included the anterior part of middle frontal gyrus and parietal lobule in both hemispheres (Fletcher et al., 1999; Seger et al., 2000), the processing of natural language typically engages the temporal lobes and the left inferior frontal gyrus, the so-called Broca’s area (Friederici et al., 2000a; Perani et al., 1998). Crucially, studies using event-related potential (ERP) measures support the view of a functional difference between nonlinguistic and linguistic grammar learning. It was shown that the pattern of brain potentials for nonlinguistic rule violations (Baldwin and Kutas, 1997) is different from those observed for the processing of violations in a linguistic artificial grammar (Friederici et al., 2002). Interestingly, the pattern of brain potentials for the latter violations is very similar to those observed for syntactic rule violations in natural language (Hahne and Friederici, 1999). The current experiment using functional magnetic resonance imaging (fMRI) set out to investigate the neural basis of artificial grammar learning by emphasizing language-like rules. Thus, in order to increase similarity between artificial grammars and natural language, the experimental design embodied an artificial language, BROCANOTO (Friederici et al., 2002), that controls for a number of variables known to influence second language acquisition: (1) a large, but poorly established, vocabulary that could delay the fast availability of syntactic word category information crucial to build up syntactic structures and (2) syntactic rules which may be different between the two languages causing interference. Therefore, the present artificial grammar (cf. Fig. 1) controlled the number of members of a given word category and the number of similar and dissimilar rules between the native and the artificial (target) grammar. Note that the syntactic structures of BROCANOTO mimic structures of natural languages (S = sentence, NP = noun phrase, VP = verb phrase). Moreover the different categories (N = noun, v = verb) are represented by different members, thus requiring an abstract representation of the word category realized as a node in the syntactic structure. BROCANOTO differs from natural languages only in that it does not include embedded sentence structures.

On the basis of previous neuroimaging studies (Strange et al., 1999, 2001) we hypothesize that learning an artificial language, in its initial phase, would be mediated by the hippocampus, with its activity decreasing as a function of increasing performance. In contrast, proficient use of the languagelike artificial grammar of BROCANOTO should
give raise to increasing activity in brain regions typically observed in normal adult language processing, like Broca’s area (Friederici et al., 2000a; Perani et al., 1998).

Methods

Subjects

Functional MRI scans were taken from 17 participants (9 male; mean age: 25 years, range: 20–29). They were all native speakers of German and right-handed and had normal or corrected-to-normal vision. Informed consent was obtained from all participants after the nature and possible risks of the experiment were explained. Three participants (two male) were excluded from all analyses due to a lack of behavioral learning effects.

Materials

The stimulus material was formed according to the artificial language BROCAN TO shown in Fig. 1. Each sentence of the artificial language, containing five to eight words, represented a subject-verb-[object] structure. The subject and the object of a sentence was a noun phrase (NP) composed of a determiner (D,d), an adjective (M), and a noun (N). The verb phrase (VP) consisted of a verb (v) and an optional adverb (m). All words forming a particular word class (nouns, verbs, etc.) were specified by a particular vowel (e.g., u = noun and e = verb). A total of 30 sentences were formulated according to these rules. Another 60 contained a severe syntactic violation. These incorrect sentences were derived by replacing tokens of one word category in a correct sentence by tokens of a different word category. Thus, three types of violations were constructed covering a wide range of possible violations. These include violations of the phrase structure, violations of the determiner-noun-agreement, and word class repetition. The number of words used in this experiment was limited to eight words in order to minimize demands on memorizing particular items.

Experimental procedure

The present experiment was similarly constructed to previous studies examining nonlinguistic artificial grammar systems (Fletcher et al., 1999; Strange et al., 2001). It comprised 15 triplets of blocks and within each triplet a fixed order of a learning block (70 s), a test block (70 s), and a sensorimotor control block (20 s) was presented visually. A brief instruction (7 s) began each block. During learning, participants viewed 10 correct sentences for 7 s each on a computer monitor and were instructed to extract the underlying grammatical rules. During test blocks, participants were presented with 10 sentences (7 s) that were either grammatical (half of the sentences) or ungrammatical. The participants task was a grammaticality judgment on each sentence presented. Visual feedback was given for each response. To control for unspecified changes of hemodynamic activity over time, blocks containing a sensorimotor control task were included after each of the test blocks. These control blocks consisted of serial visual presentation of either of two pseudowords (BRAD and DABA) which required a forced-choice button press relating to which pseudoword had been presented. Ten minutes after scanning, a transfer grammatical judgment test was performed in which 30 new sentences were presented, half followed the same grammatical rules as during learning and half were new ungrammatical sentences.

Data acquisition

Imaging was performed on a 3T Bruker Medspec 30/100 system using a standard birdcage headcoil. Procedures included collection of anatomical images using a MDEFT sequence (Ugurbil et al., 1993; Norris, 2000, data matrix 256 × 256, TE = 10ms, TR = 1.3 s). Functional images sensitive to blood-oxygen-level-dependent (BOLD) contrast were acquired using a gradient-echo EPI sequence (TE = 30 ms, flip angle = 90°, and TR = 1 s). An acquisition volume consisted of 12 axial slices (5 mm thickness, skip 2 mm) with a 3-mm² in-plane resolution. Slices were positioned to cover the medio temporal lobe up to the superior frontal sulcus. A total of 2670 functional volumes per participant were acquired in two scans synchronized with the onset of sentence presentation.

Data analysis

Data were analyzed using the general linear model employing a random effects model as implemented in the software package LIPSIA (Lohmann et al., 2001). The data were smoothed spatially with a 6-mm FWHM isotropic Gaussian kernel, and the global means were normalized by proportional scaling. We focused on learning-related changes in neuronal activity. Therefore, the time series of each condition was weighted to model a Condition × Time interaction. The model function was derived from each individual’s behavioral learning curve by fitting a logarithmic function a ln(t) + b. Our contrasts tested for greater learning effects in the learning and testing blocks versus sensorimotor control blocks. The contrasts entered into a one-sample t test across all subjects. Activations were considered significant when composed of 10 or more contiguous voxels surviving a threshold of P < 0.0001, uncorrected.

Results

Behavior

The behavioral data showed improved performance across blocks (see Fig. 2). This was indicated by an
ANOVA contrasting the performance of the first three consecutive blocks with the last three blocks \(F(1, 13) = 44.8, P < 0.0001\). Moreover, the performance of the very first block did not significantly differ from chance level (percentage correct = 52\%, t(13) < 1). The performance improvement across the experimental session could be best modeled using a logarithmic regression explaining 76\% of the experimental variance. However, participants may have transferred syntactic rules from their German mother tongue to the artificial language rather than acquiring new syntactic rules. This potential transfer was controlled for by introducing two grammatical rules (determiner-noun-agreement) into BROCAN TO that do not exist either in German or in any other natural language known to the participants (Fried erici et al., 2002). Thus, if any observed response to syntax violations rested on a rule transfer from German to BRO CANTO, these new rules should exhibit non- or slower learning. An ANOVA contrasting subjects performance for the different violation types across learning blocks revealed no distinction in learning rates as indicated by a nonsignificant Block \(\times\) Violation type interaction \(F(12, 156) = 1.04, P < 0.4\). Based on this we can conclude that the observed learning-related changes in hemodynamic activity was indeed caused by acquiring a miniature language with negligible transfer from other language. In a separate transfer test several minutes after scanning with new sentences either conforming to the learned rules or not, participants correctly classified 74\% of the sentences. This performance is significantly above chance level \(t(13) = 4.89, P < 0.001\), cf. Fig. 2 but was not significantly different from the performance level of the last learning blocks \(t(13) = 1.87, P < 0.1\), cf. Fig. 2. This provides additional evidence that the grammar of BROCAN TO was successfully acquired, that is, long-lasting representations of the underlying grammatical rules had been established.

\[y = 13.3 \ln(x) + 58.2, R^2 = 0.763\]

\[\text{Performance [\% correct]}\]

\[\text{Block} 1-3, 4-6, 7-9, 10-12, 13-15, \text{TEST}\]

**Fig. 2.** Performance (with standard error bars) across participants during learning and the subsequent transfer test. The performance measure was averaged across the three consecutive learning blocks. Language acquisition was evidenced by a significant increase in performance during the experiment indicated by an explained variance of \(R^2 = 0.76\).

**Neuroimaging**

We focused our analysis on learning-related changes during the acquisition of the artificial language. Thus, the specific contrast of interest tested was a Condition (learning versus sensorimotor control) \(\times\) Time interaction resembling a logarithmic increase or decrease of activity during the learning blocks relative to the sensorimotor control. Figure 3I shows the brain’s response and its temporal modulation while learning an artificial language. During learning, left posterior hippocampal activity [at coordinates \(-26, -32, 0\) in Talairach space (Talairach and Tournoux, 1988)] decreased relative to the recurring control task. Additionally, other brain regions also exhibited decreasing activity during learning, including the retrosplenial cortex and left anterior and posterior temporal areas (cf. Table 1). The reverse response pattern, namely significantly increased engagement during language acquisition, was found in the opercular part of the left inferior frontal gyrus, that is, Broca’s area restricted to BA44 (cf. Figure 3II, Table 1). Additional increased activity with the progression of language acquisition was also obtained in posterior association areas at the anterior part of the middle occipital sulcus in both hemispheres (BA18/19).

**Discussion**

In this study we addressed the question of the functional architecture of the acquisition of a new artificial grammar that meets the universal principles of a natural language. Our behavioral data provide evidence of transfer of classification performance to novel sentences, confirming that subjects in the present study learned an abstract linguistic grammar system on the basis of rules and not merely on the basis of superficial similarities. The imaging data during learning exhibit decreasing activity in the left posterior hippocampus, whereas the activity in the left fronto-opercular cortex increased.

The decreasing hippocampal response of the present study is consistent with the transient task-related decrease in hippocampal activation previously observed in studies on nonlinguistic grammar learning (Strange et al., 1999, 2001). In addition, probabilistic learning also produces a relative decrease in hippocampal activation (Poldrack et al., 1999). We attribute this hippocampal response pattern to similarity-based learning, which apparently governs the initial state of learning the artificial grammar and which is no longer used once the grammatical rules of the artificial language have been acquired. Supporting evidence is provided by a number of neuropsychological and neuroimaging findings of left hippocampal involvement in verbal memory processing (Helmstaedter et al., 1997; Wagner et al., 1998).
addition, relational processing which is taken to underlie pattern-based learning (Gomez and Gerken, 2000) also appears to be a function of the hippocampal system as indicated by a wide range of fMRI studies (Cohen et al., 1999; Schacter and Wagner, 1999). Furthermore, it has been suggested that relational processes based on perceptual similarity can operate in parallel to abstract rule learning (Smith et al., 1998; Strange et al., 2001). Thus, these findings are compatible with our interpretation of the present hippocampal response decreasing with time. This interpretation, however, is in apparent conflict with previous neuropsychological findings of Knowlton and Squire (1994, 1996), who reported normal performance in the artificial grammar-learning task for amnesic patients. There are several possible accounts for these patient data: Either this inconsistency may be due to the heterogeneity of the patient sample as most of the patients suffer from diencephalic damage causing amnesia or even a small amount of hippocampal functioning is sufficient to support artificial grammar learning.

Another, more likely account for these contradictory results, however, is the notion that superficial similarity between exemplars does not itself play an obligatory role in artificial grammar learning (Knowlton and Squire, 1994).

Other brain regions exhibiting decreasing activity during learning include the thalamic nuclei and the retrosplenial cortex. Both areas have a rich interconnectivity with the medial temporal lobe and together they constitute the so-called extended hippocampal diencephalic system (Aggleton and Brown, 1999). The integrity of this system is a prerequisite for successful retrieval from episodic memory. In particular, neuroimaging studies (von Zersen et al.,

Table 1
Brain areas exhibiting learning-related changes in activation

<table>
<thead>
<tr>
<th>Cortical region</th>
<th>Peak location</th>
<th>z score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Decreases in activity during learning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left posterior hippocampus</td>
<td>−26 −32 0</td>
<td>12.80</td>
</tr>
<tr>
<td>Left thalamus</td>
<td>−18 −16 1</td>
<td>12.54</td>
</tr>
<tr>
<td>Retrosplenial cortex BA 29/30</td>
<td>−8 −43 24</td>
<td>−6.40</td>
</tr>
<tr>
<td>Left midtemporal gyrus BA 22</td>
<td>−49 −34 −1</td>
<td>−6.09</td>
</tr>
<tr>
<td>Left temporal pole BA 22</td>
<td>−52 3 −4</td>
<td>−7.19</td>
</tr>
<tr>
<td><strong>Increases in activity during learning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left inferior frontal gyrus BA44</td>
<td>−40 8 18</td>
<td>9.72</td>
</tr>
<tr>
<td>Left middle occipital gyrus BA19</td>
<td>−40 −81 28</td>
<td>10.15</td>
</tr>
<tr>
<td>Right middle occipital gyrus BA19</td>
<td>25 −85 29</td>
<td>6.05</td>
</tr>
<tr>
<td>Left inferior occipital gyrus BA18</td>
<td>−29 −79 8</td>
<td>7.41</td>
</tr>
<tr>
<td>Right intraparietal gyrus BA7</td>
<td>28 −72 37</td>
<td>7.67</td>
</tr>
</tbody>
</table>

Fig. 3. (I) Brain regions in which a significant Condition × Time interaction was observed. Sagittal sections at $x = −26\text{mm}$ (left panel) and $x = −40\text{mm}$ (right panel) exhibit brain areas with changes of activity during learning relative to the recurring control blocks. Regions demonstrating decreased activity during artificial language learning included the left posterior hippocampus (A: $−26, −32, 0$) and the left ventrolateral thalamus (B: $−18, −16, 1$). Increased activity was noted in the left inferior frontal gyrus (C: $−40, 8, 18$) and the left medial occipital gyrus (D: $−40, −81, 28$). (II) Temporal changes of activation in the hippocampus (left panel) and inferior frontal gyrus (right panel) for learning periods (solid line) relative to sensorimotor control blocks (dashed line). The time course shows the response collapsed across the three consecutive learning blocks and averaged across all participants.
point to a specific role of this area in integrating various aspects of episodic information rather than in storage of episodic memories.

The lateral temporal lobes, also demonstrating decreased activity, are thought to be engaged in lexical-semantic processing (Kotz et al., 1999; Newman and Twieg, 2001). Because the vocabulary of BROCAN'TO is composed of pronounceable items that, however, do not exist either in German or any other natural language known to the participants, this could lead to an activation of the lateral temporal lobes during the initial stage of learning (Newman and Twieg, 2001). The rapid decrease of activity in these regions may be interpreted as the system’s reaction to the irrelevance of lexical-semantic aspects in the present study.

In contrast to the decreasing activity in the hippocampus, an increased engagement of the opercular part of the left inferior frontal gyrus was obtained. The involvement of prefrontal regions in abstract nonlinguistic rule learning was recently reported (Fletcher et al., 1999; Seger et al., 2000; Strange et al., 2001). However, these studies consistently found left hemispheric activation of the anterior part of the IFG (BA47), the MFG, and the frontal pole. Thus, there is a substantial difference between the cortical regions in the PFC previously found to mediate abstract rule learning and the circumscribed activation of Broca’s area observed in the present study. This might be attributed to the language-like character of the rules in our artificial grammar. Although Broca’s area is classically defined to include BA44 and BA45 the present functional specificity of BA44 can be connected to a recently described cytoarchitectonical differentiation. A morphological left-over-right asymmetry was found for BA44, but not for BA45 (Amunts et al., 1999). Broca’s area was shown to be involved in syntactic processes in a number of neuropsychological studies, in particular in that patients with lesions including Broca’s area clearly suffering from syntactic deficits (Grodzinski, 2000; Caplan and Waters, 1999). In addition, some recent brain imaging studies have demonstrated increased hemodynamic activity in Broca’s area during the processing of syntactic aspects during language comprehension (Dapretto and Bookheimer, 1999; Embick et al., 2000; Friederici et al., 2000a; Just et al., 1996; Moro et al., 2001; Tettamanti et al., 2002). Some of these fMRI studies report the activation to be focused in BA44. Direct evidence for the view that learning a new artificial grammar involves language-specific processes is provided by ERP studies. Proficient usage of the artificial language used in the present experiment elicited a biphasic ERP pattern specific to natural language processing (Friederici et al., 1996, 2002). For the early, negative portion of the biphasic ERP pattern neuronal generators in the inferior fronto-lateral cortex, that is, Broca’s area, have been reported (Friederici et al., 2000b).

In addition to our predictions increased activity in left posterior parietal and occipital areas, particularly the posterior intraparietal sulcus (IPS), was also observed. It is well established that the IPS plays an important role in a large variety of tasks with high attentional demands (Culham and Kanwisher, 2001). More recently, language-related tasks have also been reported to activate posterior parietal cortex (Gurd et al., 2002; Sohn et al., 2000). Sohn and colleagues observed increased IPS activity when switching between two verbal discrimination tasks as compared to task repetition and argued that the inferior posterior parietal cortex may be responsible for stimulus-driven task adjustment. In light of these findings it is conceivable that the increase in IPS observed in the present study might be related to adjusting the focus of attention to particular elements occurring in different positions in every trial.

Taken together, the present results point to dynamic interactions between the left hippocampal formation and left inferior frontal region during the course of the acquisition of a language-like artificial grammar. Crucially, they suggest a transition of the relative roles of different cerebral structures when a new artificial grammar is learned. As a consequence, possible implications of the present results for second language acquisition should be considered. According to a recent model of the functional architecture of first and second language (Ullman, 2001) language processing depends on temporal lobe structures as well as left inferior frontal and basal-ganglia structures. Crucially, this model posits for second language prevailing involvement of the temporal lobes in grammar processing. Previous ERP studies on second language processing had implied such involvement of posterior (and lateral) temporal areas in syntax processing. Thus, for second language learners they consistently showed an enhanced negativity over posterior brain areas, possibly reflecting demanding syntactic integration processes (Hahne and Friederici, 2001; Weber-Fox and Neville, 1996). In contrast, an early frontal effect of syntactic processing, typically observed for native speakers, was not observed for the second language learners. However, an early frontal syntax effect was reported for high proficient adult second language learners (Friederici et al., 2002). This is consistent with the present results of initial involvement of the posterior hippocampus and a shift to left inferior frontal activity associated with the proficient use of a languagelike artificial grammar.

The shift in cerebral activity, found in the present study, is also in agreement with models of memory consolidation that postulate a gradual reorganization of the neural substrates of memory representations in the course of learning (McClelland et al., 1995; Squire and Alvarez, 1995). According to these models, the learning process is accomplished by means of an interaction between the hippocampal system and the neocortex in establishing permanent cortical representations, thereby becoming independent of the hippocampal region. In the case of artificial language learning, there is a transition from general, similarity-based learning systems in the medial temporal lobes to a neural organization of language processing usually observed by lesion
studies and by functional imaging studies of the normal human brain, including the left inferior prefrontal cortex.

**Acknowledgments**

We thank Axel Mecklinger for valuable discussion and suggestions to earlier versions of the article. This work was supported by the Leibniz-Prize of the German Research Foundation awarded to Angela D. Friederici, Director at the Max-Planck-Institute of Cognitive Neuroscience, and Grant FOR448 of the German Research Foundation.

**References**


McClelland, J.L., McNaughton, B.L., O’Reilly, R.C., 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from successes and failures of connectionist models. Psychol. Rev. 7, 1569–1576.


