

The neural bases of the learning and generalization of morphological inflection



Michael Nevat^a, Michael T. Ullman^b, Zohar Eviatar^c, Tali Bitan^{a,c,*}

^a Language and Brain Plasticity lab, IIPDM, University of Haifa, Israel

^b Department of Neuroscience, Georgetown University, United States

^c Department of Psychology, University of Haifa, Israel

ARTICLE INFO

Keywords:

Inflectional morphology
Language learning
Artificial language
Type frequency
Phonological cue
fMRI

ABSTRACT

Affixal inflectional morphology has been intensively examined as a model of productive aspects of language. Nevertheless, little is known about the neurocognition of the learning and generalization of affixal inflection, or the influence of certain factors that may affect these processes. In an event-related fMRI study, we examined the neurocognition of the learning and generalization of plural inflections in an artificial language, as well as the influence of both affix type frequency (the proportion of words receiving a given affix) and affix predictability (based on phonological cues in the stem). Adult participants were trained in three sessions, and were scanned after the first and last sessions while inflecting trained and untrained words. Untrained words yielded more activation than trained words in medial frontal (including pre-SMA) and left inferior frontal cortices, which have previously shown activation in compositional grammatical processing. A reliance on phonological cues for untrained word inflection correlated positively with pre-SMA activation, but negatively with activation in the pars triangularis. Thus, pre-SMA may be involved in phonological cue-based composition, while the pars triangularis underlies alternative processes. Inflecting trained items yielded activation in the caudate head bilaterally, only in the first session, consistent with a role for procedural memory in learning grammatical regularities. The medial frontal and left inferior regions activated by untrained items were also activated by trained items, but more weakly than untrained items, with weakest activation for trained-items taking the high-frequency affix. This suggests less involvement of compositional processes for inflecting trained than untrained items, and least of all for trained inflected forms with high-frequency affixes, consistent with the storage of such forms (e.g., in declarative memory). Overall, the findings further elucidate the neural bases of the learning and generalization of affixal morphology, and the roles of affix type frequency and affix phonological predictability in these processes. Moreover, the results support and further specify the declarative/procedural model, in particular in adult language learning.

1. Introduction

Affixal inflectional morphology has often been used as a model for investigating the learning and processing of productive aspects of both first and second language, and their neural bases. Inflectional affixes within a given morphological system tend to vary in the extent of their applicability, based on various factors such as affix type frequency (the proportion of words receiving a given affix; Croft (2007)) and affix predictability (the degree to which the affix can be predicted from phonological or other cues in the stem). Indeed, as we shall see, behavioral evidence suggests that both affix type frequency and affix predictability seem to modulate both the learning and generalization of affixal inflection. However, we are not aware of any previous studies

investigating the *neural* bases of the influence of these factors on affixal inflection, or even the neural substrates of learning and generalizing affixal inflection more generally. The current study was designed to address these gaps, with possible broader relevance to other productive aspects of language. Specifically, the aim of this behavioral and fMRI study was to elucidate the neurocognition of the acquisition and generalization of affixal inflection in adults learning an artificial language, in a multi-session training paradigm, while probing the influence of affix type frequency and affix predictability.

* Corresponding author at: Department of Psychology, University of Haifa, 199 Abba Khoushy Ave., Haifa, Israel.
E-mail address: tbitan@research.haifa.ac.il (T. Bitan).

<http://dx.doi.org/10.1016/j.neuropsychologia.2016.08.026>

Received 29 January 2016; Received in revised form 15 August 2016; Accepted 25 August 2016

Available online 26 August 2016

0028-3932/ © 2016 Elsevier Ltd. All rights reserved.

1.1. Behavioral evidence regarding the influence of affix type frequency and affix predictability on affixal inflection

1.1.1. Evidence from natural languages

The influence of both affix type frequency and affix predictability on affixal inflection has been examined in natural language. To date such behavioral research has focused on first language, with little work on second language. Additionally, most such studies have investigated the effects of generalizing inflectional affixes (to novel or irregular forms), with little research probing the effect of these variables on the learning or processing of existing inflected forms.

At least in first language, higher affix type frequency seems to be associated with a greater tendency to generalize inflectional affixes. (We are aware of no work on affix type frequency in second language.) For example, *Dabrowska and Szczerbiński (2006)* found that 2 and 3 year old children's application of inflectional affixes to nonce words in Polish was positively correlated with the inflectional affixes' frequencies. In other studies, overgeneralization (over-regularization) rates of inflectional affixes to irregulars (e.g., 'goed') seem to correlate with the inflection's affix type frequency, with fairly high rates for the (high affix frequency) regular '-ed' past tense and '-s' plural inflections in English (*Maslen et al., 2004*), but low rates for the (low affix frequency) '-s' plural inflections in German (*Köpcke, 1998*).

When words with common semantic or phonological characteristics take the same inflectional affix, these characteristics can act as cues to the affix. The degree to which cues can reliably predict inflectional affixes can vary. The predictability of an inflectional affix, given a cue, can be defined as the proportion of words with the cue that take the affix out of the total number of words containing the cue. In some languages, semantic and phonological cues are correlated. For example, gender serves as a cue for the selection of plural inflectional affixes for Hebrew nouns (-ot vs. -im), and is itself at least partly predictable based on word-final phonemes (*Berent et al., 1999; Ravid et al., 2008*). In other languages (e.g., plural inflections of German nouns; *Laaha, 2011*) phonological cues are not correlated with semantic cues, and both types of cues may help predict the correct inflection. Phonological cue predictability has been found to correlate positively with both the learning and generalization of inflectional affixes, in both first and second language. *Laaha (2011)* found that native German speaking children perform better at producing existing inflected forms that have more predictable plural affixes, suggesting that these forms were better learned. In native speakers greater phonological cue predictability also appears to be associated with a higher generalization rates of inflectional affixes, both to novel forms (*Albright and Hayes, 2003*) and to irregulars, in the form of over-regularizations (*Hartshorne and Ullman, 2006*). Note that the examination of phonological predictability for stem-changing irregulars (*Pinker, 1991; Pinker and Ullman, 2002*) is not discussed here, as we focus on affixal inflection. Finally, higher phonological predictability has been found to improve both the learning and generalization of affixal inflection in second language (*Kempe and Brooks, 2008*).

1.1.2. Evidence from artificial languages

Although most previous work on the influence of affix type frequency and affix predictability on morphology has, not surprisingly, examined natural language, research has begun to turn to artificial languages to examine these issues. Artificial language paradigms are particularly well suited for examining learning and generalization because one can tightly control the amount and type of language exposure, such as manipulating factors of interest in the input. Artificial linguistic paradigms have the added advantage that, likely because they are small, they can generally be learned to reasonably high proficiency over the course of hours to days, thereby enabling the longitudinal examination of language learning and generalization.

Hence, despite concerns regarding their ecological validity because they do not reflect the full complexity of natural languages, artificial

languages have been widely used in the investigation of both vocabulary (*Tamminen et al., 2012; Davis et al., 2009*) and grammar (*Ellis and Schmidt, 1997; Merks et al., 2011; Morgan-Short et al., 2012a, 2012b*). Importantly, performance at artificial language learning has been found to correlate positively with natural second language learning (*Ettlinger et al., 2016*), and training on an artificial language can result in native-like brain activity patterns (*Morgan-Short et al., 2012a, 2012b*). Thus, results from artificial languages show a likelihood of generalizability to natural languages.

Note that although researchers have used artificial language paradigms as models of first language acquisition (e.g., *Karuzza et al., 2013*), in the present study we interpret the learning and generalization of the artificial language as a model of second language learning (e.g., *Morgan-Short et al., 2012a, 2012b*), since in this study learning occurs in adulthood (when participants have already learned at least their first language), and moreover, as in a second language, the artificial language involved learning inflections for familiar items (e.g., apple; see Methods).

We are aware of three artificial language studies investigating the influence of affix type frequency on the learning or generalization of affixal inflection. In two studies, *Ellis and Schmidt (1997, 1998)* found that higher affix type frequency facilitates the acquisition of trained inflected forms in an artificial language. Similarly, *Bybee and Newman (1995)* observed that higher affix type frequency improved the generalization of affixes to untrained words.

Additionally, in a recent artificial language study we examined the effects of both affix phonological predictability and affix type frequency on the learning and generalization of affixal inflection (*Nevat et al., under review*). In this purely behavioral study, we used an artificial language paradigm similar (but not identical) to the one examined in the present study. Three groups of adult participants were trained on plural inflectional suffixes in the artificial language, with an orthogonal manipulation of suffix type frequency and phonological predictability across groups. The results indicated that participants inflected trained words with high-frequency suffixes more accurately than those with medium- and low-frequency suffixes (with the worst performance on those with medium-frequency suffixes). Moreover, for untrained words participants relied on the predictability of rime cues when selecting the affix, a reliance which increased with exposure to the language. These findings reveal the importance of both suffix type frequency and suffix phonological predictability in the learning and generalization of affixal morphological inflection in an artificial language learned as an adult.

1.2. Relevant theoretical and empirical neurocognitive research

Although behavioral studies are beginning to elucidate the influence of affix type frequency and affix predictability on affixal inflection, as mentioned above we are aware of no prior research on the neural bases of the effects of these factors on affixal inflection, nor more generally on the functional neuroanatomy of the learning and generalization of affixal inflection. Nevertheless, prior theoretical and empirical neurocognitive research on other aspects of language provides a foundation on which to examine these issues.

1.2.1. A neurocognitive theoretical account: the declarative/procedural model

A number of neurocognitive models have been proposed to explain the processes involved in learning and processing a second language, and how these may differ, overlap, or interact with those underlying first language (e.g., *Abutalebi, 2008; Clahsen and Felser, 2006; Hernandez et al., 2005; Paradis, 1994; Ullman, 2015*). The model that appears to make the most specific neuroanatomical predictions for grammar learning in a second language, and which provides our primary predictions, is the Declarative/Procedural (DP) model (*Ullman, 2001a, 2001b, 2004, 2005, 2015, 2016*). We therefore focus on this model here. Note that this does not imply that the other models

are incorrect, or that our results are inconsistent with these models; however, such comparisons are not the focus of this study.

According to the DP model, language learning, storage, and use depend heavily on two general-purpose learning and memory systems, declarative memory and procedural memory (see following paragraphs for specifics on these systems). The model posits that idiosyncratic knowledge, including of simple words and irregular morphological forms, is always learned and stored in declarative memory. Rule-governed grammatical knowledge, in contrast, can be learned and processed in either system. In both first and second language aspects of grammar are generally learned initially in declarative memory (e.g., as chunks or explicit rules), since this system learns faster than procedural memory. However, gradually procedural memory learns the underlying rules, which eventually become automatized. Although this occurs in both first and second language, there should be an increased and longer lasting dependence on declarative memory for grammar in later-learned second language, since learning in declarative memory improves over the course of childhood, while learning (or consolidation) in procedural memory may attenuate. Indeed, adults learning a second language may never proceduralize aspects of their grammar as much as first language learners, perhaps even after years of exposure.

The procedural memory system, which is rooted in frontal/basal-ganglia circuits, underlies the implicit learning of a wide range of motor and cognitive skills (Eichenbaum, 2003; Gabrieli, 1998; Squire, 2004; Ullman, 2004, 2016). Note that we use the term procedural memory to refer to a particular brain system and its characteristics, rather than implicit memory more generally, which is how some researchers use the term. Skilled performance requires the extraction of recurring elements from a series of separate events (Squire, 2004), and is therefore gradually acquired. In the network of brain structures underlying procedural memory, the basal ganglia play a critical role in the learning and consolidation of motor and cognitive skills, whereas frontal regions may be more important for processing skills after they have been automatized. Within the basal ganglia, the caudate nucleus (and the anterior putamen) may be especially important for skill acquisition (Ashby et al., 2007; Doyon and Benali, 2005); the head of the caudate may be particularly important (Ullman, 2004, 2016). Implicated frontal regions include premotor and related cortex, including the pars opercularis of the inferior frontal gyrus (Brodmann's area (BA) 44), lateral premotor cortex (BA 6, in the precentral gyrus and extending more anteriorly), and the supplementary motor area (SMA; note that pre-SMA is the anterior portion of SMA).

The declarative memory system has traditionally been defined as the brain system that underlies explicit knowledge (i.e., knowledge that can be brought to conscious awareness). This brain system, which is well studied in both humans and non-human animals, is rooted in the hippocampus and other medial temporal lobe structures (Eichenbaum, 2003; Gabrieli, 1998; Squire, 2004; Ullman, 2004, 2016). These structures are critical for the learning and consolidation (stabilization) of new knowledge, which however eventually relies largely on neocortical regions (Davis and Gaskell, 2009). Other brain structures also play a role, such as anterior and ventral regions of the inferior frontal gyrus (BA 45 and 47, including pars triangularis of the inferior frontal gyrus), which seem to be involved in aspects of recall (Ullman, 2004, 2016). Declarative memory, which evidence now suggests underlies implicit as well as explicit knowledge, may be specialized for learning and representing idiosyncratic (non-derivable) information and arbitrary associations; indeed, this system may be necessary for learning such information and associations.

The two memory systems also interact. Of particular interest here, there appears to be a negative relation between the two systems, which we refer to as the seesaw effect (Ullman, 2004, 2016). For example, the dysfunction of one system can result in the enhancement of the other. Moreover, learning in declarative memory may inhibit learning in

procedural memory (Ullman, 2004, 2016; Poldrack and Packard, 2003).

1.2.2. Relevant neurocognitive evidence

Two broad lines of prior empirical research are particularly relevant to the examination of the neural bases of learning and generalizing affixal inflectional morphology. First, a substantial number of neuroimaging studies have probed affixal morphology. However, these have focused on the processing (rather than the learning or generalization) of regular (affixal) inflection, primarily in adults, and mainly in first language (e.g., Beretta et al., 2003; Desai et al., 2006; Sahin et al., 2006; Tyler et al., 2005), though also in second language (Pliatsikas et al., 2014). We are aware of no neuroimaging studies of the learning or generalization of affixal morphology in artificial language paradigms. Of interest here, although results vary somewhat across studies, regular inflection has been associated with inferior frontal regions, especially left pars opercularis (BA 44) (Beretta et al., 2003; Desai et al., 2006; Pliatsikas et al., 2014; Sahin et al., 2006; Tyler et al., 2005), but also SMA (Desai et al., 2006; Sahin et al., 2006), as well as other regions. For a recent neuroanatomical meta-analysis of regular and irregular inflection that specifically implicates left pars opercularis in regular morphology in first language, see Ullman et al., (in preparation).

Second, as mentioned just above, neuroimaging studies of second language, as well as of artificial grammar or artificial language learning paradigms, have largely ignored morphology, instead focusing on other aspects of grammar, including syntax. These studies relevant to grammar learning have often implicated the left inferior frontal cortex, including the pars opercularis (Musso et al., 2003; Opitz and Friederici, 2003; Newman-Norlund et al., 2006; Zaccarella and Friederici, 2015). Studies of second language grammar and of artificial grammar learning have also demonstrated involvement of the caudate nuclei, in particular the caudate head (Ullman, 2015, 2016). For example, the caudate nuclei have been implicated in the processing of grammatical (as compared to non-grammatical) sequences in neuroimaging studies of artificial grammar learning (Lieberman et al., 2004; Forkstam et al., 2006). Both inferior frontal cortex and the basal ganglia (especially the caudate head) have also been implicated in grammar in second language in a neuroanatomical meta-analysis (Tagarelli et al., in preparation).

1.3. The present study

The present event-related fMRI study used an artificial language to examine the neurocognition of the learning and generalization of affixal inflectional morphology over the course of acquisition, and how this activity is affected by two factors: suffix type frequency and suffix phonological predictability. In this artificial language, plural inflection, the target inflection of the study, was formed by adding suffixes to stems. Participants encountered five different suffixes during training. These differed in affixal type frequency and were generally (but not always) determinable by phonological cues embedded in the stems.

Based on the predictions of the DP model, we expected to find activation for the learning and generalization of affixal morphology in procedural memory as well as declarative memory brain structures. In particular, caudate nucleus activation was expected, especially the caudate head, while learning the affixal inflectional system (and thus at early stages of acquisition), whereas activation of premotor and related regions (e.g., pars opercularis and (pre-)SMA) was expected for the processing of the regularities, especially in their generalization to untrained forms. The processing of trained items was expected to activate these regions to a lesser extent, instead perhaps relying on declarative memory structures, such as pars triangularis (for the recall of the stored information). In addition, medial temporal lobe structures might be expected during the learning of these forms. Although evidence from previous behavioral studies suggested a likelihood of both affix type frequency and affix phonological predictability influen-

cing the learning and generalization of affixal inflection, no specific anatomical predictions were made for these factors due to the absence of prior neurocognitive studies.

2. Methods

2.1. Participants

26 participants (15 women), ages 20–47 (mean 26.25 years, SD 6.09) were recruited for this study among university and college students in Israel. Of these, 4 participants (3 women) could not be scanned due to claustrophobia, and technical malfunctions prevented the completion of data collection from 5 additional participants (2 women). Therefore, the results reported here are based on data from 17 participants (10 women). All participants were native Hebrew speakers and spoke at least one other language (English) as an additional language. All participants reported being right-handed, had no known psychiatric, neurodevelopmental, or neurological disorders, and had normal or corrected hearing and vision.

2.2. Materials

The *trained* items consisted of 48 nouns, which were aurally presented in the context of the artificial language. All items consisted of two syllables (always CVCVC) in their singular form (the stem). Plural forms were always obtained by applying one of 5 possible (VC) suffixes to the stem. The high-frequency suffix was applied to half of the items (24 items), the medium frequency suffix was applied to one quarter of the items (12), and three low-frequency suffixes were each applied to one twelfth (4) of the items. Pairings of stems and suffixes were generally determined by the stems' rimes; see Table 1. For example, words ending with /oz/ took the high-frequency suffix ('-an'); thus the plural for 'tuvoz' was 'tuvozan'. However, some trained items did not follow these rules. These "exception" words, which contained "inconsistent rime cues", rhymed with other words in the trained item list, i.e., those that (generally) took a different suffix. For example, the stem 'shalod', which received the high-frequency suffix '-an' rhymed with stems that received the medium-frequency suffix ('-esh'), such as 'napod' and 'resod'. This was done with the purpose of introducing ambiguous cues into the lists, thereby affecting the predictability of inflectional affixes given these cues. Participants were not informed of any of the patterns underlying stem-suffix pairings.

The structure of this inflection system resembles inflectional systems in natural languages in a number of respects. The existence

of multiple suffixes of varying frequencies is found in the plural inflectional system of nouns in German (Laaha, 2011; Marcus et al., 1995). Phonological cues at word-ending positions that predict inflectional suffixes are also found in natural languages, including in the plural inflection of nouns in both Hebrew (Berent et al., 1999; Ravid et al., 2008) and German (Laaha, 2011). Importantly, however, the use of an artificial language enabled full control over other properties, even beyond the careful manipulation of affix type frequency and affix predictability. Hence, the first three phonemes of each word (CVC) were selected to ensure they would not provide any cue to the inflection. Moreover, the stimuli were designed so that the grammatical gender of the Hebrew translation of the noun would not provide any cue to the inflection.

In addition to the 48 items that participants were trained on, they were tested (at the end of the first and third sessions; see below) on the production of plural forms of 36 *untrained* items that contained rime cues. Different lists of 36 untrained items were presented in each of the two tests. Of the 36 untrained words in each list, 12 contained rime-cues that predicted the high-frequency suffix (i.e. they rhymed with trained items whose stems contained consistent cues and took the high-frequency suffix), 12 contained rime-cues that predicted the medium-frequency suffix, and 12 contained rime-cues that predicted the low-frequency suffix (see Supplementary Material for the lists of untrained items).

2.3. Procedure

The experiment, which encompassed procedures both outside the scanner and inside the scanner, took place over the course of three sessions, one every 3–4 days (see Fig. 1). This time frame has been shown in previous studies to be sufficient for consolidation and lexicalization of newly learnt words (Davis et al., 2009).

2.3.1. Outside the scanner

2.3.1.1. Instruction block. The first session of the experiment began with an instruction block (see Fig. 1), which exposed participants to all of the trained items in both their singular and plural forms, together with their meanings. Each of the 48 training items was presented once. For the design of each trial in the instruction block, see Fig. 2a. Each trial (corresponding to one item) began with the presentation of a fixation cross. When the participant pressed the space bar the singular form was presented aurally together with an image of a real object (e.g., an apple, pen, carrot) on the computer screen. Participants were

Table 1

List of trained items. Presented by suffix frequency (high, medium, low), and the predictability of each suffix given the cue in the stem.

High-frequency inflectional suffix Applied to 24 trained items Suffix: '-an'	Medium-frequency inflectional suffix Applied to 12 trained items Suffix: '-esh'	Low-frequency inflectional suffixes Applied to 12 trained items: 3 suffixes x 4 items each Suffixes: '-ev', '-ak', '-ur'																																										
<p>Items containing "consistent rime cues" (predictability: 1.0)</p> <p>3 "families" of 6 items each:</p> <table border="0"> <tr> <td>nifoz</td> <td>nishig</td> <td>tizul</td> </tr> <tr> <td>tuvoz</td> <td>posig</td> <td>shuzul</td> </tr> <tr> <td>kufozl</td> <td>bolig</td> <td>mupul</td> </tr> <tr> <td>lalo</td> <td>dedjig</td> <td>suful</td> </tr> <tr> <td>refoz</td> <td>rekig</td> <td>tedjul</td> </tr> <tr> <td>gishoz</td> <td>givig</td> <td>bikul</td> </tr> </table>	nifoz	nishig	tizul	tuvoz	posig	shuzul	kufozl	bolig	mupul	lalo	dedjig	suful	refoz	rekig	tedjul	gishoz	givig	bikul	<p>Items containing "consistent rime cues" (predictability: 0.8)</p> <p>"families" of 4 items each:</p> <table border="0"> <tr> <td>napod</td> <td>paniv</td> <td>koshun</td> </tr> <tr> <td>nezod</td> <td>tepiv</td> <td>rosun</td> </tr> <tr> <td>resod</td> <td>lekiv</td> <td>ligun</td> </tr> <tr> <td>moshod</td> <td>sibiv</td> <td>batun</td> </tr> </table>	napod	paniv	koshun	nezod	tepiv	rosun	resod	lekiv	ligun	moshod	sibiv	batun	<p>Items containing "consistent rime cues" (predictability: 1.0)</p> <p>No two items' stems rhyme with each other:</p> <table border="0"> <tr> <td>Suffix: '-ev'</td> <td>dipem</td> <td>tegas</td> <td>sapor</td> </tr> <tr> <td>Suffix: '-ak'</td> <td>lidek</td> <td>mikal</td> <td>nerud</td> </tr> <tr> <td>Suffix: '-ur'</td> <td>getav</td> <td>nised</td> <td>rinit</td> </tr> </table>	Suffix: '-ev'	dipem	tegas	sapor	Suffix: '-ak'	lidek	mikal	nerud	Suffix: '-ur'	getav	nised	rinit
nifoz	nishig	tizul																																										
tuvoz	posig	shuzul																																										
kufozl	bolig	mupul																																										
lalo	dedjig	suful																																										
refoz	rekig	tedjul																																										
gishoz	givig	bikul																																										
napod	paniv	koshun																																										
nezod	tepiv	rosun																																										
resod	lekiv	ligun																																										
moshod	sibiv	batun																																										
Suffix: '-ev'	dipem	tegas	sapor																																									
Suffix: '-ak'	lidek	mikal	nerud																																									
Suffix: '-ur'	getav	nised	rinit																																									
<p>Items containing "inconsistent rime cues"</p> <ul style="list-style-type: none"> ○ 3 items whose stems rhyme with stems that take the medium frequency inflection (shalod, gukiv, gitun). (predictability: 0.2) ○ 3 items whose stems rhyme with stems taking low frequency inflections (kunus, gomil pakom). (predictability: 0.5) 		<p>Items containing "inconsistent rime cues" (each stem rhymes with one item taking the high-frequency suffix) (predictability: 0.5)</p> <ul style="list-style-type: none"> ○ meshus (suffix: '-ev') ○ shibil (suffix: '-ak') ○ zufom (suffix: '-ur') 																																										

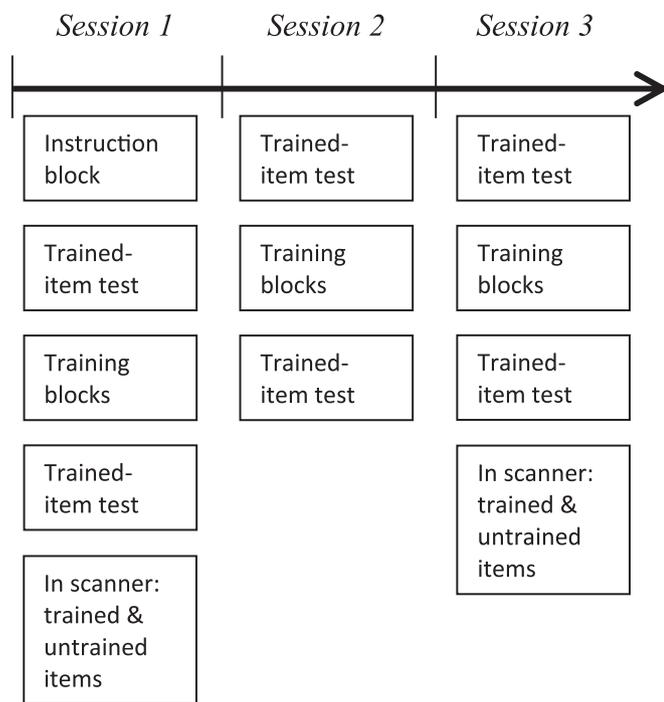


Fig. 1. Overall design of the experiment. The instruction block, training blocks, and trained-item tests were given outside the scanner. During the instruction block, given at the beginning of Session 1, participants heard the singular and plural form of all to-be-trained items, and repeated the plural form. During the training blocks (five per session) participants produced inflected forms, given their stems, and then heard the correct inflected form as feedback. The trained-item tests, given before and after training in each session, consisted of correct/incorrect judgment of inflected forms. Inside the scanner, participants both repeated and inflected both trained and untrained items.

informed at the beginning of this instruction block that these were the objects the items referred to. The images were included in order to increase the resemblance of the artificial language to natural languages, and because semantic information has been shown to be important for the lexicalization of newly learned words (Leach and Samuel, 2007; Merx et al., 2011). The use of familiar objects (which correspond to existing words in the participants' first language) makes this artificial language paradigm more similar to learning a second than a first language. The singular form was followed by a visual cue consisting of two asterisks (**) in the center of the screen for one second, indicating that the plural form of the word would soon be presented. The plural form was then presented aurally, followed by the presentation of a question mark in the center of the screen, indicating that participants were to repeat the plural form they had just heard. This cue remained on the screen for a maximal duration of four seconds, or until a vocal response was detected.

2.3.1.2. Trained-item tests. Trained-item tests requesting the judgment of correctly and incorrectly inflected plural forms were presented both before and after each series of training blocks (see Fig. 1). Thus, there were two trained-item tests in each of the three sessions. Each of the 48 trained items was tested once in each trained-item test block. The design of the trials in the trained-item tests is presented in Fig. 2b. In each trial an asterisk was first presented as a visual cue for 500 ms, followed by an aural presentation of the singular form of a trained item. No images were presented during this task. The singular form was followed by a visual cue consisting of two asterisks (**), which was presented for one second, followed by an aurally presented plural form of the same word. In each trained-item test block, half of the 48 plural forms presented were correct, and half were incorrect. Incorrect inflections were created by adding one of the other

suffixes to the stem. Across the two trained-item test blocks within each session (i.e., before and after each series of training blocks), each word was presented once with a correct inflection and once with an incorrect inflection. The order of the presentation of correct and incorrect inflections was counterbalanced across participants. Across sessions each participant was presented with all different incorrectly affixed forms. The presentation of the plural form was followed by a question mark in the center of the screen, indicating that a response was required. Participants were instructed to press "1" if the plural form was correct, and to press "2" if not. They were given 3 s to respond.

2.3.1.3. Training. In each of the three sessions, participants also underwent training (Fig. 1), in which they practiced the production of inflected forms on the 48 trained items. During training, each trial (see Fig. 2c) began with the presentation of a fixation cross until the space bar was pressed, followed by the aural presentation of the singular form of a trained item together with a visual presentation an image of the object the trained word referred to. As soon as the aural presentation of the word was completed, a question mark was displayed to prompt participants to pronounce the plural form of the word they had just heard. Once recording of the response terminated, or three seconds elapsed without a vocal response, the correct plural form of the word was presented aurally, as feedback to the participant. The training session consisted of five blocks, in each of which each trained item was presented once. The order of items within each training block was randomized.

2.3.2. Inside the scanner

In the 1st and 3rd sessions, event-related fMRI scans were performed after the second trained-item test (see Fig. 1). During these scans participants both repeated and inflected both trained and untrained items. Trained and untrained items were presented in separate runs. For both trained and untrained items, in each run trials were divided into alternating "blocks" in which participants were instructed to generate plural forms of words presented to them ("inflection"), or to repeat the presented words ("repetition"). Each such "block" consisted of 8 trials. Fig. 3 presents the design of trials in these tasks. Each trial began with the presentation of a fixation point at the center of the screen for 500 ms. A word was then presented aurally, and an image of an object was presented on the screen for 1 s. Stimulus presentation was followed by a 3–5 s interval ("covert phase") during which participants were instructed to think of either the correct inflected form (in inflection blocks) or the singular form (in repetition blocks) according to the color of the cues presented on the screen (white for inflection blocks, red for repetition blocks). A question mark was then presented on the screen for 2.5 s, at which point participants had been asked to orally produce the same form they had thought of during the "covert phase". No feedback was provided. The trial terminated with a 3–5 s interval ("inter-trial interval", or ITI). Durations of the covert phase and the ITI were jittered (3–5 s) so as to enable independent estimation of the hemodynamic response to the covert and oral response phases, while keeping trial durations as short as possible ("rapid presentation"; e.g., Serences, 2004). Words were presented in a randomized order.

In each of the two scanner sessions, participants were presented with all 36 untrained items, as well as 36 of the 48 trained items (see the Materials section). The 36 trained items comprised 12 items taking the high-frequency suffix (only those with consistent rime cues), 12 taking the medium-frequency suffix (all of which have consistent rime cues), and 12 taking the low-frequency suffixes (nine of which contain consistent rime cues, three of which contain inconsistent rime cues). Thus, all items taking medium- and low-frequency suffixes were included. Of the items taking the high-frequency suffix, only items

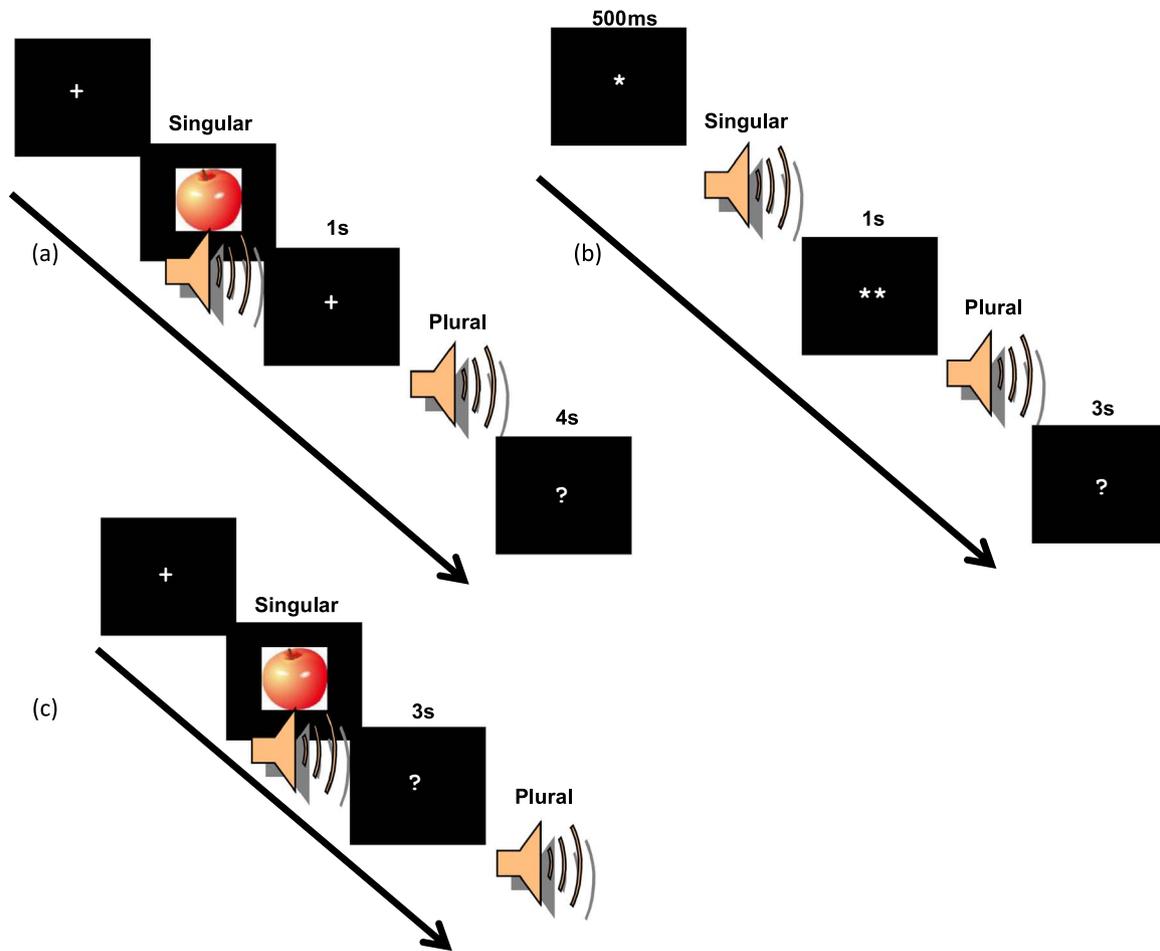


Fig. 2. Outside the scanner. Design of trials in (a) instruction block, (b) trained-item tests, (c) training.

with consistent rime cues were included; items with inconsistent rime cues taking the high-frequency suffix were therefore omitted, as were two items of each rime family with consistent rime cues taking the high-frequency suffix. The fact that high-frequency suffixed items with inconsistent rime cues were not presented in the test should not have biased the phonological predictability of affixes because, unlike during training, no feedback was provided.

Similarly, the 36 *untrained* items included 12 words containing rime-cues that predicted the high-frequency suffix (i.e. they rhymed with the stems of trained items that contained consistent cues and took the high-frequency suffix), 12 containing rime-cues that consistently predicted the medium-frequency suffix, and 12 containing rime-cues that consistently predicted the low-frequency suffix (see

[Supplementary Material](#) for the lists of untrained words).

For both the trained and untrained items each stem was presented twice during inflection blocks, and twice during repetition blocks. Thus altogether 144 trained and 144 untrained items were presented (72 inflection trials and 72 repetition trials), with a mean duration of 12 s (range: 10–14 s). Trained and untrained items were each presented in three 9:36 min runs.

2.4. MRI data acquisition

Images were acquired using a 3-Tesla GE Signa scanner, with an 8-channel head coil. Head movement was minimized using a cushion inserted between the head and the head coil. Auditory stimuli were

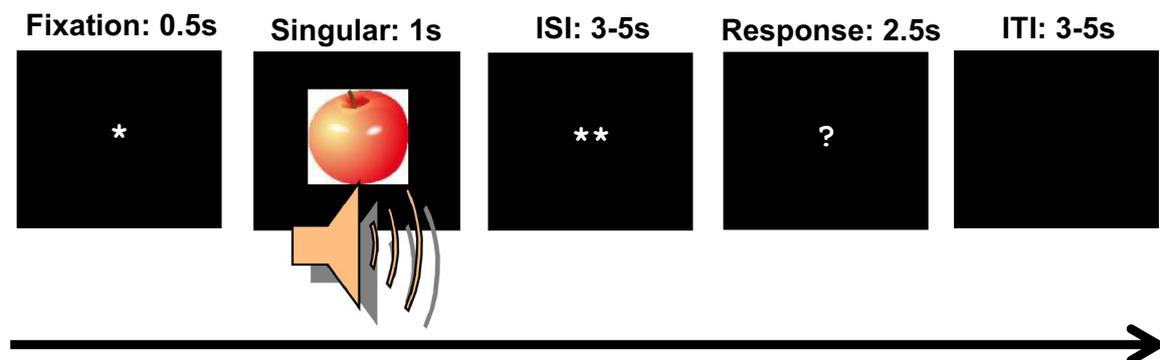


Fig. 3. Inside the scanner. Design of trials for repetition and inflection. All cues appeared in white (as presented here) during inflection trials, and in red during repetition trials. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

presented through pneumatic earphones, and visual stimuli were projected onto a screen, and viewed through a mirror attached to the head coil. Participants' responses were recorded using an optical microphone (FOMRI-II- Optoacoustics, Mazor, Israel). The BOLD functional images were acquired using the EPI method. The following parameters were used for scanning: TE =35 ms, flip angle=90°, matrix size=64×64, field of view=24 cm, slice thickness=3.2 mm, gap=0.8 mm, number of slices=33, TR=2000 msec. 288 images were acquired during each run. In addition, structural T1-weighted 3-D images were acquired (SPGR, TR=7.948 msec, TE=3.036 msec, flip angle=20°, matrix size=256×256, field of view=24 cm, slice thickness=1 mm, number of slices=168) using an identical orientation as the functional images.

2.5. Analyses

2.5.1. Behavioral

2.5.1.1. Trained items. The learning of affixal inflection in trained items was examined on judgment data from the trained-item tests, which were analyzed with two-way repeated measures ANOVAs. The (within-subjects) factors were suffix frequency (high, medium or low) and “test number”, an index taking values from 1 to 6 indicating the point at which each of the six trained-item tests was administered (1: 1st session, before training; 2: 1st session, after training; 3: 2nd session, before training, etc.). The dependent variables were accuracy and reaction time. In all ANOVAs the Greenhouse–Geisser correction was applied for sphericity values lower than 0.75 and the Huynh–Feldt correction was applied for sphericity values greater than 0.75 (see Field, 2005).

2.5.1.2. Untrained items and their reliance on phonological cues. In order to examine whether participants had acquired knowledge of the phonological cues embedded in the trained items, and the extent to which behavioral performance and brain activation were related to this knowledge, we determined the extent to which participants' actual responses on untrained words matched the responses that would be predicted based on these cues. For this purpose, we calculated the predictability of each suffix given each rime cue, using the following formula: $\text{Predictability}(\text{suffix}, \text{cue}) = \frac{p(\text{suffix}|\text{cue})}{p(\text{cue})} = \frac{n_{\text{suffix} \cap \text{cue}}}{n_{\text{cue}}}$; where $n_{\text{suffix} \cap \text{cue}}$ represents the number of items that contain the cue and receive a specific suffix; and n_{cue} represents the total number of items that contain the cue. Thus, this is simply a measure of how strongly a given suffix is predicted by the rime cue in the trained stimuli. See Table 1 for the values of predictability thusly computed. The list of trained items was designed so as to minimize correlations between the factors of suffix phonological predictability and suffix frequency.

The suffix with the highest predictability for a given cue is referred to here as the “optimal” response. Participants' actual responses for each untrained item were scored (1 or 0) according to whether they matched the optimal response. In cases for which there was more than one optimal response (i.e., two suffixes had the same highest predictability) the score was divided in half. The total scores (across all items) were summed for each participant in each session, and then divided by the total number of trials in which participants were asked to inflect untrained words, in order to compute proportions. The optimal response score is very close to an accuracy measure (and in fact it is identical to “accuracy” for items with rimes that provide unambiguous cues). However, its advantage is in its applicability to items with rimes that provided ambiguous cues (for which there was more than one “correct” response, so optimality is defined by the suffix with the highest predictability). Proportions of optimal responses thus computed were entered into single-sample *t*-tests in order to determine

whether they were greater than 20%, which is the chance level expected if suffixes were selected at random (100% divided by 5 suffixes). To determine whether the reliance on phonological cues had increased with training, the proportions of optimal responses in the 1st and 3rd sessions were also compared, using a paired-samples *t*-test.

2.5.2. fMRI images

fMRI data were analyzed using the Statistical Parametric Mapping toolbox for Matlab (SPM8: Wellcome Trust Center for Neuroimaging, University College London, www.fil.ion.ucl.ac.uk/spm). The images were spatially realigned to the first volume in each run. The average maximal translation per run was 1.22 mm (range: 0.3–5.77), and the average maximal rotation was 1.28° (range: 0.22–5.78). Images containing a displacement of 4 mm or greater, or a rotation of 4° or greater, were omitted from analyses. In total, 520 images (or 0.93% of images acquired) were thus discarded. These discarded images were taken during 9 different tasks/sessions, and acquired from 5 of the participants. Sinc interpolation was used to minimize timing errors between slices (Henson et al., 1999). The functional images were coregistered with the anatomical image, and normalized to the standard T1 template volume (MNI). The data were then smoothed with a 5 mm isotropic Gaussian kernel. An event-related design was used to estimate hemodynamic responses. A high-pass filter with a cutoff period of 128 s was applied. Movement parameters calculated during realignment were included as regressors. Hemodynamic responses were modeled using 6 finite impulse responses (FIR), each lasting 2 s, with onsets of 0–10 s after event onset.

Two types of models were specified at the first level analysis. The first, referred to as the “global” model, was applied separately to the trained and untrained items in each of the two scanning sessions. Thus, a total of four such models were specified per participant. Each of these models included five conditions at the first (participant) level: “covert inflections”, “overt inflections”, “invalid inflections” (i.e., “invented” suffixes and missing responses), “covert repetitions”, and “overt repetitions”. These models were used for the whole-brain and ROI analyses. At the second (group) level, analyses were based on the comparison between activity in response to the covert production of inflected forms and activity for covert repetitions (excluding items whose overt production included invalid suffixes). The covert phase, which is the time (3–5 s) from the presentation of the singular form until the cue prompting the production of the overt response (i.e., either the inflected form or the repeated singular form) is the period of interest to us here, since this is when participants are assumed to process the stimuli and select the inflectional affix in the inflected condition. We were not interested in the overt phase in which the oral response is executed, which may contain speech movement artifacts, and therefore did not analyze fMRI data from this phase.

The second type of model (“frequency analysis”) was applied only to correct responses for trained items in the 1st session, and was used to identify effects of frequency on brain activity. Six conditions were defined: “high-frequency covert inflections”, “medium-frequency covert inflections”, “low-frequency covert inflections”, “overt inflections”, “covert repetitions”, and “overt repetitions”. Here, as above, overt responses (overt inflections and overt repetitions) were not of interest, and were included in the model only in order to reduce errors in estimations of activation related to covert processing; activation from overt responses was therefore not included in group analyses. Data from this model were used in ROI analyses. This model could not be applied to data from untrained words because the number of trials in which medium- and low-frequency suffixes were “optimally” applied was insufficient. Additionally, there was no point in applying this model to data from trained-item tests in the 3rd session, as data from the global model indicated that activation during inflections in these tests did not differ from baseline in most ROIs (see below).

2.5.2.1. Whole-brain analysis. The whole brain analysis was designed to reveal brain regions involved in covert affix inflection (as compared to covert repetition), over trained and untrained words, across both sessions. To this end, contrast images in which “covert inflections” were compared to “covert repetitions” in the first level (participant) analysis were included in the second level (group) analysis. The group analysis combined both trained and untrained items and both sessions, in a “flexible factorial” design, resulting in four conditions: “trained items, 1st session”, “untrained items, 1st session”, “trained items, 3rd session”, and “untrained items, 3rd session”. A map of all regions showing significant activation across trained and untrained items and across sessions in this model served as the basis for the identification of regions of interest (ROIs).

2.5.2.2. ROI analyses. The ROI analyses were designed to reveal finer-grained effects of brain activity during affix inflection, related to trained vs. untrained, 1st vs. 3rd session, affix frequency, and affix phonological predictability. These analyses examined the network of regions that were involved in covert inflections (compared to covert repetitions), as determined by the whole-brain group analysis (see just above). This approach avoids the danger of circularity between selection and selective analyses (Kriegeskorte et al., 2009). This is because the contrast used for the selection of ROIs (all inflections vs. baseline) is orthogonal to the contrasts used for the selective analyses in ROIs (trained vs. untrained or 1st vs. 3rd sessions), and all conditions include an equal number of trials.

The statistical map obtained by comparing inflection with repetition was assessed for cluster-wise significance, with a cluster-defining threshold of uncorrected $p=0.001$. At this threshold the uncorrected critical cluster size was 150 voxels. Changes in signal intensity during covert inflections and covert repetitions in the first level models in functionally defined masks were extracted using the Marseille Boîte À Région d'Intérêt (MarsBaR, v.0.43- see Brett et al. (2002)) toolbox for SPM. MarsBaR calculates “percent signal change” by comparing activation during a condition of interest to average activation within the same ROI throughout the session (<http://marsbar.sourceforge.net/faq.html>). Data thus extracted were analyzed using IBM SPSS Statistics software (v. 19). Differences in signal change (% signal change during inflections – % signal change during repetitions) were calculated. These were entered into repeated measures ANOVAs, one per cluster, with session (1 vs. 3) and trained vs. untrained items as within-subject variables. In all ANOVAs the Greenhouse–Geisser correction was applied for sphericity values lower than 0.75 and the Huynh–Feldt correction was applied for sphericity values greater than 0.75 (see Field, 2005). Clusters exhibiting significant effects of trained/untrained and/or session number were divided into regions of interest using neuroanatomical masks from the Automated Anatomical Labeling atlas (AAL), as done in Nevat et al. (2014). This was done in order to differentiate among sub-regions within clusters, some of which were of considerable volume and crossed different anatomical regions (see Table 3). Only clusters with at least a 150 voxel overlap with the functional mask were included.

In order to examine the extent to which brain activation during the inflection of untrained items was related to knowledge of phonological cues, we first identified ROIs that were more active during untrained compared to trained items in the analysis described just above. We then tested the correlation between percent signal change in these regions during inflection (vs. repetition) of the untrained items and proportions of “optimal” responses (see above), which serve as an indicator of knowledge of phonological cues. We focused on the activation during the inflection of untrained words because participants do not have “word-specific” knowledge of these words, and therefore only for untrained words a high proportion of optimal responses can

serve as an indication of reliance on phonological cues.

In order to examine effects of suffix frequency on brain activation we first identified clusters that showed greater activation for inflections compared to repetitions in trained items in the first session, based on the “global” model above. We then extracted the magnitudes of signal change in each of these regions for each affix frequency category (high, medium, and low), as well as magnitudes of signal change during repetition trials, from the “frequency analysis” model. Differences in signal change were calculated and compared using one-way repeated-measures ANOVAs (one for each cluster). Clusters exhibiting significant effects of session were divided into smaller ROIs for further analysis, as described above.

3. Results

3.1. Analyses of behavioral measures

3.1.1. Trained items

As described in the Methods, the effects of experience and affix frequency on learning affix inflection in the trained items was examined with two-way 6 (test number: 1–6) by 3 (affix frequency: high, medium, low) ANOVAs. These revealed, first of all, that participants' performance on both accuracy and reaction times improved with experience (main effect of test number: $F_{(3,07, 43,00)}=43.58$, $p < 0.001$, and $F_{(2,22, 31,01)}=28.70$, $p < 0.001$, for accuracy and reaction times respectively). Additionally, participants judged items with high-frequency affixes more accurately than those taking both medium- and low-frequency affixes, which did not differ from each other (main effect of frequency $F_{(1,79, 25,07)}=8.95$, $p < 0.01$, followed by pairwise comparisons between levels of frequency, significant $ps < 0.01$). Reaction times on the judgment task were slower overall for items with medium frequency affixes than those with both the high- and low-frequency inflections, which did not differ from each other (main effect of affix frequency $F_{(2,28)}=10.24$, $p < 0.01$, followed by pairwise comparisons, significant $ps < 0.01$). See Table 2 and Fig. 4. No other effects were significant.

3.1.2. Untrained items

As described above, participants' reliance on phonological cues in

Table 2
Accuracy and reaction times (RT) for trained-item tests.

Test Number	Affix Frequency	Accuracy (%)		RT (ms)	
		Mean	Std. Deviation	Mean	Std. Deviation
1	High	69.72	8.25	779.97	266.13
	Medium	69.44	16.86	851.06	234.65
	Low	63.89	10.76	839.50	342.32
2	High	89.17	6.26	526.55	236.19
	Medium	72.22	15.32	700.32	339.06
	Low	78.89	12.55	568.13	246.79
3	High	84.56	9.52	453.70	168.28
	Medium	71.08	18.89	553.92	206.33
	Low	81.86	14.50	500.78	202.29
4	High	93.14	7.06	332.66	129.46
	Medium	80.39	16.12	448.50	229.53
	Low	88.24	13.20	356.42	151.82
5	High	94.36	4.15	412.78	143.33
	Medium	83.82	14.27	418.44	234.69
	Low	85.29	9.56	401.57	138.89
6	High	97.06	3.54	366.46	131.27
	Medium	89.71	9.56	405.08	113.58
	Low	92.65	11.37	343.55	179.09

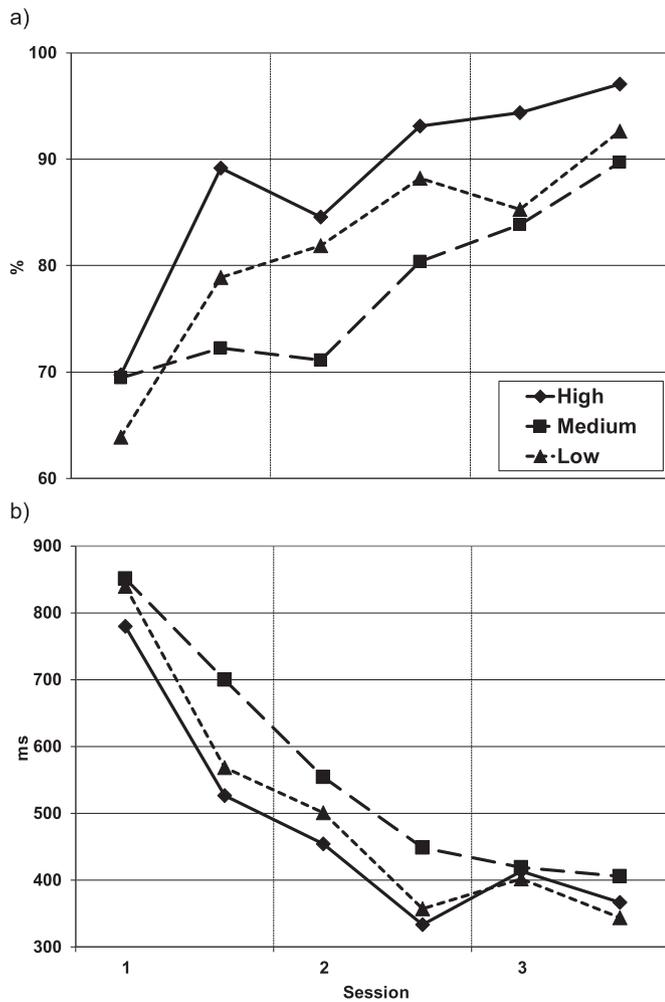


Fig. 4. Learning curves for trained items in terms of a) accuracy and b) RT. The two points in each session reflect performance in the trained-item tests (performed outside the scanner) before and after training in that session. High, Medium, and Low refer respectively to the high-, medium-, and low-frequency affixes.

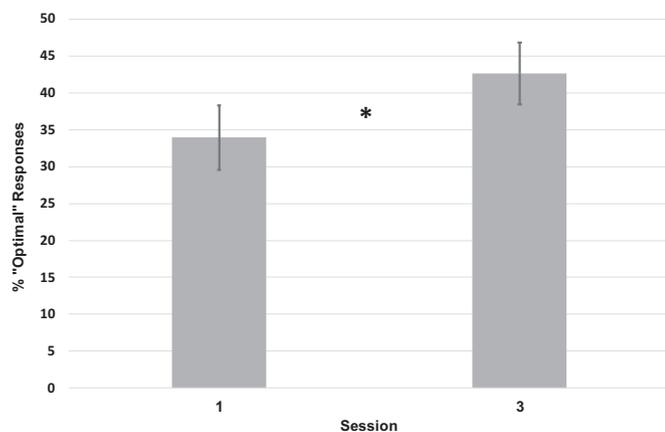


Fig. 5. "Optimal" responses by session. Error bars indicate limits of 95% confidence intervals.

their selection of inflectional affixes for untrained items was measured in terms of proportions of "optimal" responses (during overt production in the scanner), where the term "optimal" refers to the selection of the affix with the highest predictability, based on the relation between phonological cues and affixes in the trained items. The proportions of optimal responses were compared to chance levels (20%), separately for the 1st and 3rd sessions, using single-sample *t*-tests. These

indicated that proportions of optimal responses for untrained items were already produced above chance by the end of the 1st session ($t_{(16)}=6.73, p < 0.001$), and remained greater than chance at the end of the 3rd session ($t_{(16)}=11.51, p < 0.001$). Additionally, results of a paired *t*-test showed that the proportions of optimal responses had increased from the 1st to the 3rd session ($t_{(16)}=3.16, p < 0.01$), indicating that reliance on rime cues in the inflection of untrained items had increased with training. Fig. 5 presents the proportions of optimal responses in each of the sessions.

3.2. fMRI

3.2.1. Whole-brain analysis

Table 3 and Fig. 6 show the regions that were more active during covert inflection than during covert repetition, across trained and untrained items and both sessions (1st and 3rd), as obtained from group analysis based on the "global" models. As described in the Methods, we report only clusters equal to or greater than 150 voxels (which correspond to a cluster level threshold of $p_{\text{uncorrected}}=0.05$, using a cluster-defining threshold of $p_{\text{uncorrected}}=0.001$). This comparison yielded activation in six significant clusters: medial frontal regions, left inferior frontal regions, left occipito-parietal cortex, a cluster encompassing occipital cortex bilaterally, and both the left and right caudate nuclei (caudate head in both cases). These clusters were then further analyzed as regions of interest (ROIs). Each caudate nucleus was analyzed separately.

3.2.2 ROI analyses

3.2.2.1. Effects of trained/untrained items and 1st/3rd sessions. Differences between activation during covert inflection and covert repetition, based on data extracted from the "global" models described above, were entered into two-way ANOVAs, in separate analyses for each of the six clusters identified in the whole-brain analysis, in order to examine effects of trained vs. untrained items and 1st vs. 3rd session. A main effect of trained vs. untrained items, with stronger activation for untrained items, was found in the medial frontal ($F_{(1,16)}=7.19, p < 0.05$) and left inferior frontal ($F_{(1,16)}=7.63, p < 0.05$) clusters. A main effect of session was also significant in the left inferior frontal cluster ($F_{(1,16)}=5.88, p < 0.05$), and was marginally significant in the medial frontal cluster ($F_{(1,16)}=3.50, p=0.08$), in both cases due to a decrease in activation from the 1st session to the 3rd. There were no trained/untrained item by 1st/3rd session interactions in these two clusters, and no effects at all in the other four clusters.

In order to identify the specific anatomical regions within the medial frontal and left inferior frontal clusters that showed untrained/trained and session effects, separate analyses were carried out on ROIs obtained from conjunctions between these two clusters in the global analysis and neuroanatomical masks (see Methods). This resulted in five ROIs, corresponding to (portions of) left precentral gyrus, left inferior frontal gyrus pars opercularis, left inferior frontal gyrus pars triangularis, medial frontal gyrus, and "pre-SMA" (the anterior part of SMA).

Repeated measures ANOVAs with trained/untrained items and 1st/3rd sessions as the within-subject factors yielded significant main effects of trained/untrained, with greater activation for the untrained items, in pre-SMA ($F_{(1,14)}=14.32, p < 0.01$), medial frontal gyrus ($F_{(1,14)}=5.97, p < 0.05$), and left pars triangularis ($F_{(1,14)}=12.39, p < 0.01$). Marginally significant effects, again with greater activation for untrained than trained items, were also found in the left precentral gyrus and left pars opercularis ($F_{(1,14)}=3.35, p=0.089$ and $F_{(1,14)}=3.30, p=0.091$, respectively). See Fig. 7. Interestingly, effects of session were not significant (or marginally significant) in any of the three ROIs examined in the left inferior frontal cortex, despite the significant effect

Table 3

Results from the whole-brain analysis: Regions of activation during inflection (compared to repetition), across trained and untrained items and both sessions. Clusters are ordered by peak intensity (z scores). Regions with clusters significant at the threshold of $p_{FWE} < 0.05$ (extent ≥ 343) are displayed in bold (all but the right caudate head), as are the coordinates of peaks significant at the threshold of $p_{FWE} < 0.05$ at the voxel level (all but the peak of the left occipito-parietal cluster). BA: Brodmann's area. AAL: Automated Anatomical Labeling atlas. SMA: supplementary motor area. NA: Not applicable. FWE: family-wise error correction.

Region	BA	Z Score	Voxels	Strongest Peak			
				x	y	z	Location (AAL)
Bilateral occipital cortex (Lingual Gyrus/Calcarine Fissure/Inferior Occipital Gyrus/Middle Occipital Gyrus/ Fusiform Gyrus/ Cerebellum)	17/18/19	5.40	4586	29	-82	0	Fusiform gyrus
Left caudate nucleus (head)	NA	4.93	343	-11	12	-2	
Medial frontal cortex (SMA/ Medial Frontal Gyrus/Cingulate Gyrus)	6/32	4.86	666	-5	14	52	SMA
Right caudate nucleus (head)	NA	4.68	153	12	14	-2	
Left inferior frontal cortex (Inferior Frontal Gyrus/ Precentral Gyrus/Postcentral Gyrus)	44/45/6	4.61	1430	-42	6	32	Precentral gyrus
Left occipito-parietal cortex (Superior Occipital Gyrus/Middle Occipital Gyrus/Precuneus)	7/31	3.87	361	-20	-67	28	Superior occipital gyrus

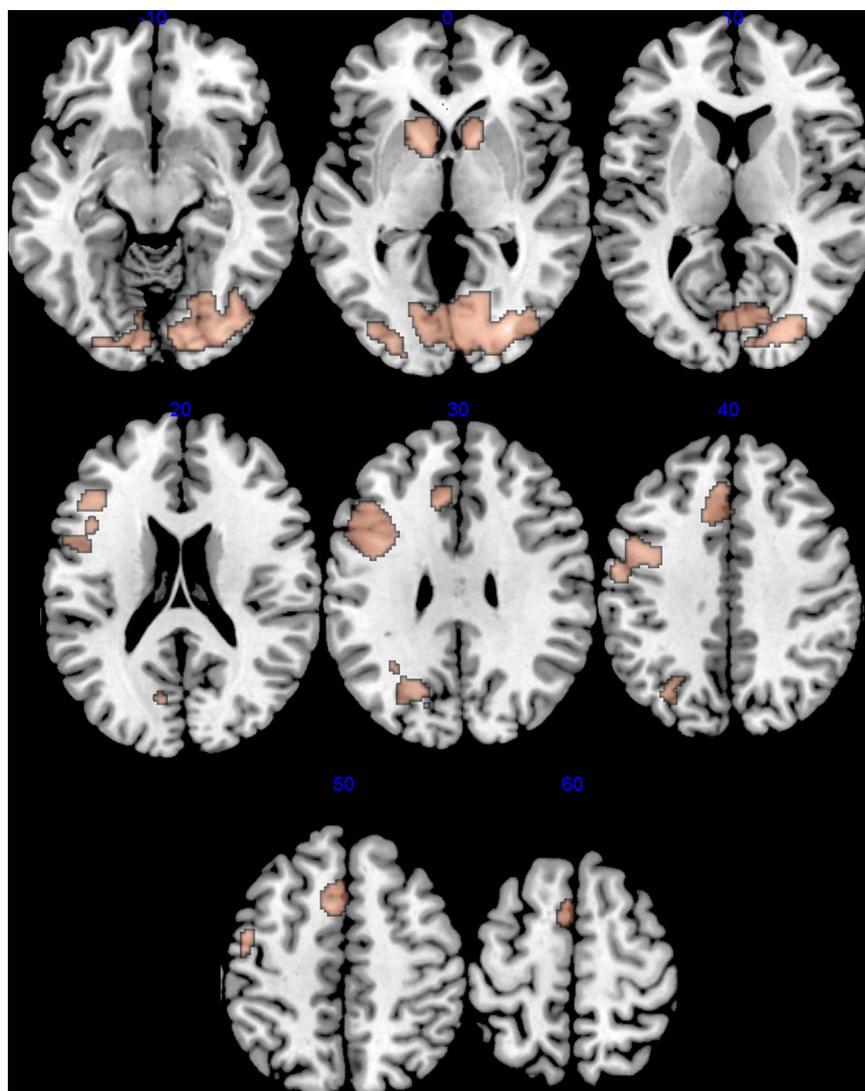


Fig. 6. Activation during covert inflection as compared to covert repetition, across trained and untrained items in both sessions.

of session at the level of the cluster. (Effects of session were not examined in the two medial frontal ROIs since the effect of session was only marginally significant at the cluster level; see above).

As discussed in the Introduction, we predicted the involvement of the basal ganglia, in particular the caudate, especially the caudate head, on the basis of the DP model and previous findings (e.g., Lieberman et al., 2004; Forkstam et al., 2006). Indeed, as we have seen in the global analyses, both the left and right caudate heads were significantly active across trained and untrained items, and across both sessions. We

expected this activation to be stronger in earlier phases, due to the apparent role of the basal ganglia in procedural learning rather than in processing the eventually automatized procedures (see Introduction). Thus, although no significant effects of session or trained/untrained items were found in the caudate, these predictions led us to examine activation in the left and right caudate head for inflection and repetition (both vs. an implicit baseline) for trained and untrained items in the 1st and 3rd sessions, each tested separately compared to 0, using single-sample *t*-tests. The results indicated that both the left and

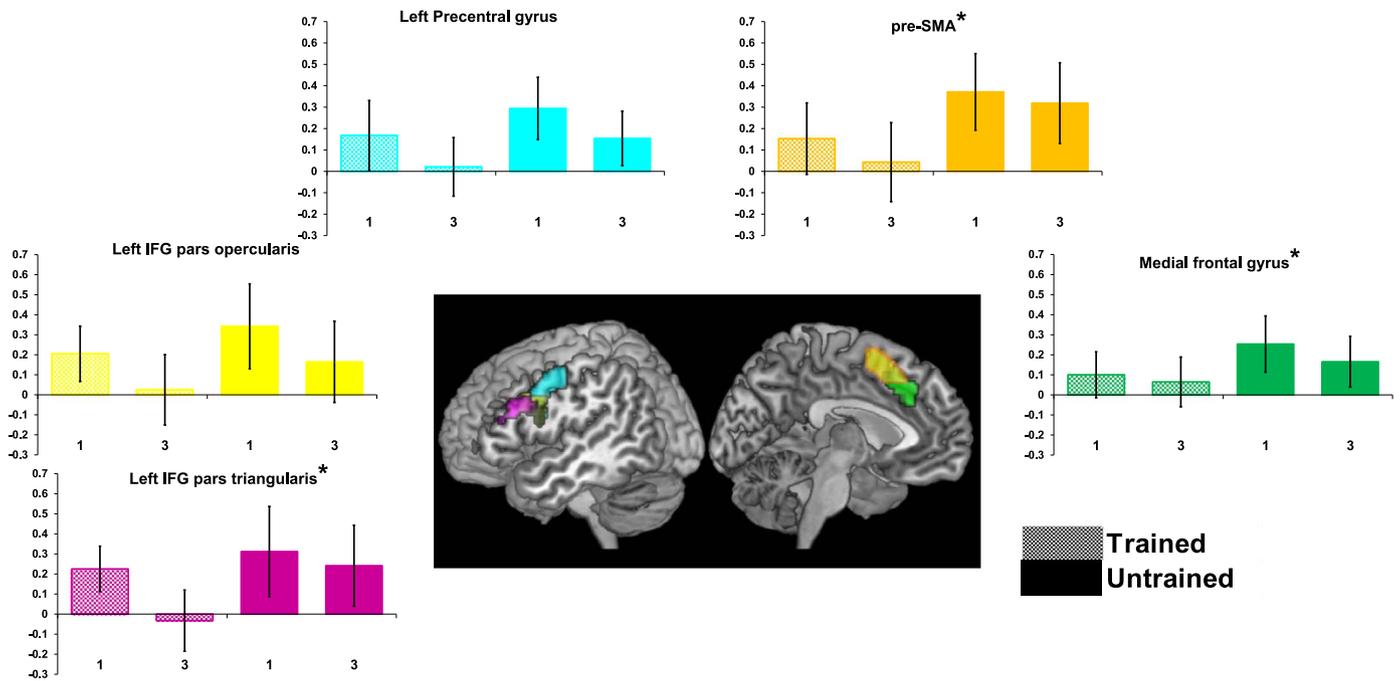


Fig. 7. Activation for covert inflection (vs. repetition) in ROIs that showed a main effect (i.e., over both sessions) of greater activation for untrained than trained items. This was significant (indicated by asterisks) for pre-SMA, medial frontal gyrus, and left pars triangularis, and marginally significant for left pars opercularis and left precentral gyrus (see text). Activation for trained and untrained items are shown separately for each session in the bar graphs. The bars represent, from left to right: trained items, 1st session; trained items, 3rd session; untrained items, 1st session; untrained items, 3rd session. Error bars indicate limits of 95% confidence interval.

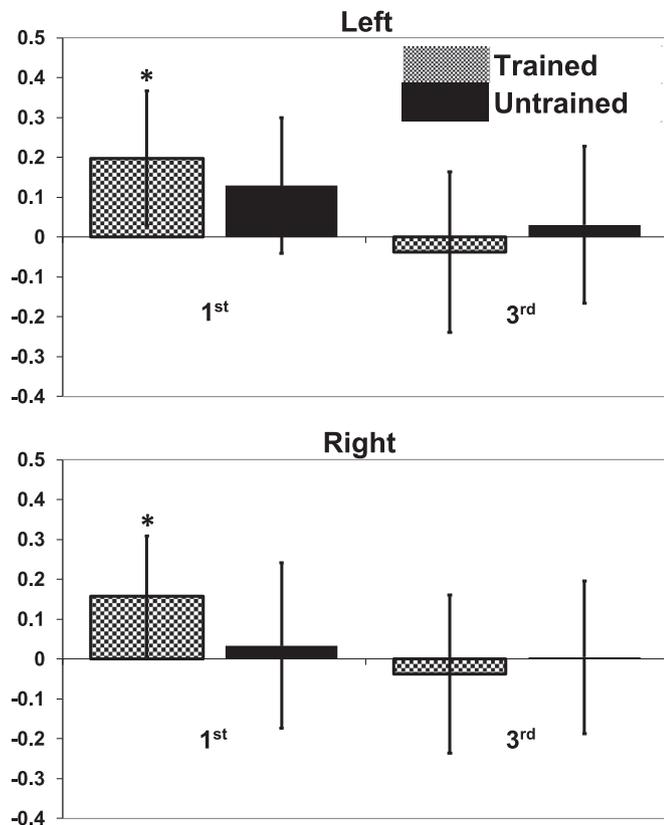


Fig. 8. Activation during inflection (vs. an implicit baseline) in the left and right caudate heads. In each case, the bars represent, from left to right: trained items, 1st session; untrained items, 1st session; trained items, 3rd session; untrained items, 3rd session. Error bars indicate limits of 95% confidence intervals.

right caudate nuclei were active only during inflection (not repetition) for trained items (not untrained items) in the 1st session (not the 3rd session): single-sample *t*-tests: $t_{(14)}=2.52, p < 0.05$ and $t_{(14)}=2.22, p <$

0.05 for left and right caudate nuclei, respectively. Fig. 8 presents the left and right caudate only for inflection, for trained and untrained items, for the two sessions.

3.2.2.2. Effects of affix phonological predictability. In order to examine whether a reliance on phonological cues is associated with brain activation during the affixation of untrained items, correlations between proportions of “optimal” responses and activation in the ROI analyses were calculated for untrained items (see Methods). Because optimal responses are indicative of reliance on phonological cues particularly for the inflection of untrained items (rather than for trained items) these correlations were calculated only for regions that were significantly more active for untrained than trained items (see above): pre-SMA, medial frontal gyrus, and left pars triangularis. We examined correlations between activation in these three regions in each of the two sessions with the measures of optimal responses in each of the two sessions, yielding a total of six correlations for each measure of optimal responses (see Table 4).

Results indicated that activation in pre-SMA during the 1st session correlated positively the proportion of optimal responses in the 1st session ($r=0.64, p < 0.05$, corrected for six comparisons); see Table 4. Proportions of optimal responses in the 1st session also correlated

Table 4
Correlations between activation for inflecting untrained words and “optimal” responses.

		“Optimal” responses	
		Session 1	Session 3
pre-SMA	Session 1	0.64*	-0.15
	Session 3	0.50	0.28
Medial frontal	Session 1	0.49	-0.22
	Session 3	0.53	0.14
Left triangularis	Session 1	0.07	-0.62*
	Session 3	-0.03	0.06

* $p < 0.05$, corrected for multiple comparisons.

positively with activation in pre-SMA in the 3rd session, and with medial frontal cortex in both sessions, though these results, which were significant prior to correction for multiple comparisons, did not survive correction (pre-SMA 3rd session: $r=0.50$, $p=0.25$ after correction; medial frontal 1st session: $r=0.49$, $p=0.27$ after correction; medial frontal 3rd session: $r=0.53$, $p=0.16$ after correction). In contrast, activation in the left pars triangularis in the 1st session correlated *negatively* with the proportion of optimal responses in the 3rd session (Table 4, $r=-0.62$, $p < 0.05$, after correction for multiple comparisons). That is, higher activation in the left pars triangularis in initial stages of training predicted less reliance on phonological cues by the end of training (i.e., those participants with more such activation in the first session showed less reliance on phonological cues in the third session). No other correlations were significant.

In order to test whether the findings in pre-SMA and left pars triangularis were independent of each other, or whether they resulted from a correlation between the two regions, we calculated the partial correlation between activation in pre-SMA and proportions of optimal responses in the 1st session, controlling for activation in left pars triangularis in the same session. We also calculated the correlation between activation in left pars triangularis in the 1st session and proportions of optimal responses in the 3rd session, controlling for activation in pre-SMA in the 1st session. These partial correlations were no weaker, and actually slightly stronger ($r=0.72$ and $r=-0.65$, respectively, $ps < 0.01$) than the original correlations, despite the fact that activation in pre-SMA and left pars triangularis in the 1st session were correlated with each other ($r=0.54$, $p < 0.05$).

3.2.2.3. Effects of affix type frequency. Analyses of the effects of affix frequency on brain activation were based on the “frequency analysis” models, as described in the Methods. As explained above (see Methods), this analysis could only be carried out for trained items, because inflections of untrained items yielded few optimal responses with medium- and low-frequency suffixes in either session, yielding low statistical power. Thus only on the trained items, we first identified clusters showing greater activation for covert inflection compared to covert repetition, based on the “global” model, separately in each session. In these analyses, five clusters (all but the occipito-parietal cluster) showed above-threshold activation for trained items in the 1st session (medial frontal: $t(16)=2.53$; left inferior frontal: $t(16)=2.56$; occipital cortex: $t(16)=2.59$; left caudate: $t(16)=2.61$; right caudate: $t(16)=2.87$; $ps < 0.05$), while no cluster showed above-threshold activation in the 3rd session. We therefore performed further analyses of the effect of frequency (low, medium, high) on activation, on ROIs for each of the five clusters that showed significant activation on trained items in the 1st session (see Methods).

Of these five clusters, a significant main effect of frequency was found only in the left inferior frontal cluster ($F_{(2,28)}=5.21$, $p < 0.05$), with marginally significant effects in the medial frontal region ($F_{(2,28)}=3.22$, $p=0.055$), and the left caudate nucleus ($F_{(1.87,26.16)}=2.92$, $p=0.075$). In all three cases the lowest activation (see Fig. 9) was found for high-frequency inflections (which showed significantly or marginally significantly lower activation than medium-frequency inflections, as well as low-frequency inflections in some cases; marginally significant differences are not shown in Fig. 9), whereas there were similar levels of activation in medium- and low-frequency inflections. The cluster in left inferior frontal regions was subsequently divided into three ROIs, as described above (i.e., overlapping with pars triangularis, pars opercularis, precentral gyrus). Significant effects of frequency were found in all three ROIs (pars triangularis: $F_{(2,28)}=4.57$, $p < 0.05$; pars opercularis: $F_{(2,28)}=4.31$, $p < 0.05$; precentral gyrus: $F_{(2,28)}=4.03$, $p < 0.05$). Again, in all three regions activation was lowest for high-frequency inflections, and similar between medium- and low-frequency inflections (Fig. 9). In two of the regions (precentral gyrus and pars triangularis) activation for low-

frequency inflections was also greater than activation for high-frequency inflections.

4. Discussion

This study examined brain activation associated with the learning and generalization of affix inflectional morphology. Participants were trained on plural inflectional suffixes of nouns in an artificial language, in which the rimes of words provided probabilistic cues to their suffix. The type frequency of the suffixes varied (i.e., some suffixes were applied to more words than others), and were classified as high-, medium- or low-frequency, according to the number of nouns that took each suffix.

Behaviorally, our results revealed that performance on trained items not only improved over the course of training but was also modulated by affix frequency. In particular, trained items with high-frequency affixes yielded the highest accuracy. Performance on untrained items indicated an increase in reliance on phonological cues over the course of training, as indicated by an increase in “optimal” responses.

The fMRI analyses revealed that the covert production of inflected forms (compared to covert repetition), across trained and untrained items in both sessions, yielded activation in six clusters: medial frontal and left inferior frontal regions, the heads of the left and right caudate nuclei, left occipito-parietal cortex, and an extensive cluster encompassing occipital cortex bilaterally. There was also a decrease in activation between the first and third sessions, across trained and untrained items, consistent with the observed improvements in performance. The following sections discuss inflections of trained and untrained words separately, as revealed in further analyses of each of these regions.

4.1. Affixal inflection of untrained words

Participants showed greater activation during the inflection of untrained than trained items, over both sessions, in the medial and left inferior frontal clusters. Region of interest (ROI) analyses in anatomically defined subregions of these clusters revealed greater activation for untrained than trained items in all subregions. This reached significance in the pre-SMA, medial frontal gyrus, and left inferior frontal gyrus pars triangularis, and showed marginal significance in the left inferior frontal gyrus pars opercularis and the left precentral gyrus. Moreover, across participants, activation in pre-SMA during the inflection of untrained items in the first session was positively correlated with a measure of reliance on rime cues (proportion of “optimal” responses) in this session. In contrast, activation in the left pars triangularis in the first session *negatively* predicted reliance on rime cues in the third session.

In contrast to trained items, inflected forms of untrained (new) words could not have been simply stored as whole forms. Therefore, greater activation in the medial and inferior frontal regions for untrained compared to trained items suggests that these regions are involved in processes underlying productive affixal morphology. Previous studies have found these regions, including the pars opercularis, precentral gyrus, pars triangularis, and (pre-)SMA, show activation for grammar, in syntax and/or productive affixal (regular) morphology, in first and/or second language (e.g., Ruschmeyer et al., 2005; Wartenburger et al., 2003; Vannest et al., 2005; Heim et al., 2005; Newman et al., 2001; Beretta et al., 2003; Tatsuno and Sakai, 2005; Desai et al., 2006; Sahin et al., 2006; Dhond et al., 2003; Tyler et al., 2005; Oh et al., 2011; Friederici, 2011). The finding that syntax as well as morphology has been found to lead to activation in these regions bolsters the view that the underlying processes are compositional. Moreover, the results are consistent with the claim of the DP model that grammatical rule-governed composition, including in productive affixal morphology, generally relies on the procedural

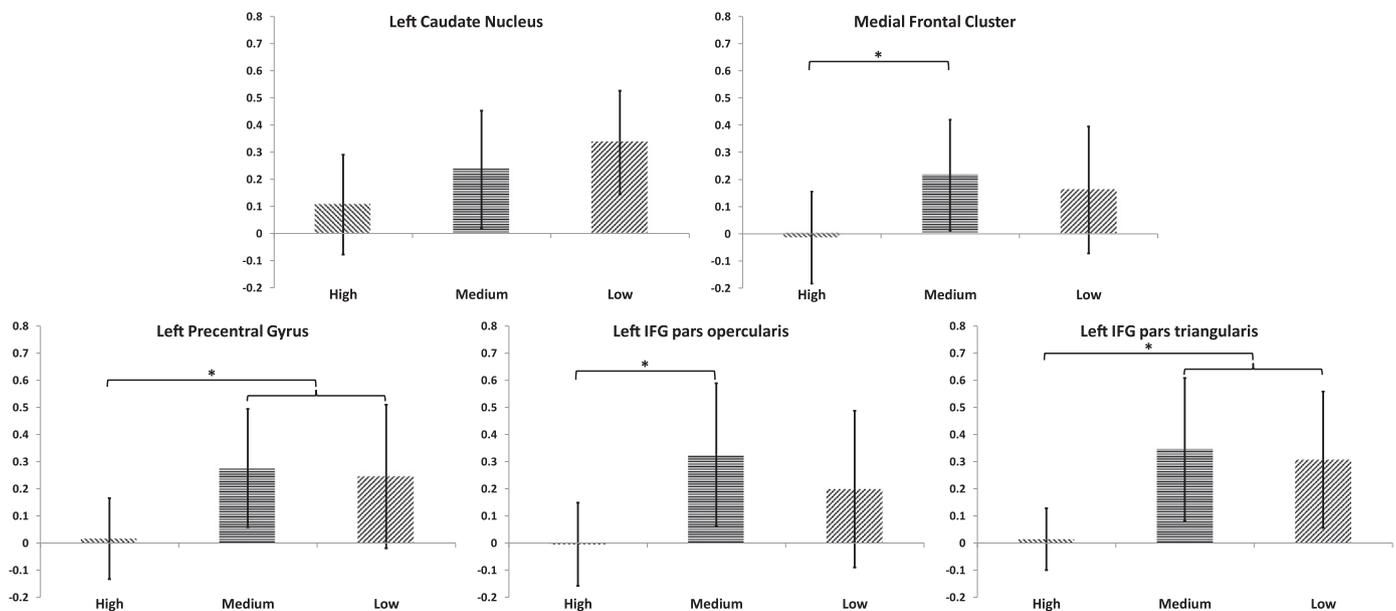


Fig. 9. Activation during inflection (vs. repetition) for trained items in the 1st session, by suffix type frequency. Error bars indicate limits of 95% confidence intervals. IFG: inferior frontal gyrus.

memory system, in both first and second language – specifically on the premotor and related regions that seem to be involved in processing learned skills in this system, in particular BA 44 (pars opercularis) and premotor cortex, including BA 6 in the precentral gyrus and (pre-) SMA. The fact that the activation pattern held across both sessions suggests that these neurocognitive processes can come into play quite early in the process of learning affixal inflection, and that they apply more to productively-computed (untrained) than already-learned (trained) forms.

Interestingly, a functional distinction has been proposed between different portions of inferior frontal cortex and nearby regions, with pars opercularis and premotor regions posited to play important roles in procedural memory-based compositional processes, while pars triangularis may be more important in the recall of information from declarative memory (Ullman, 2004, 2006, 2016). This distinction suggests the possibility that posterior inferior frontal regions (i.e., pars opercularis and precentral gyrus) and pars triangularis may reflect different processes in the production of untrained inflected forms. On this view, the posterior regions underlie their composition in procedural memory, while pars triangularis subserves the recall of relevant knowledge from declarative memory. For example, greater pars triangularis activation for untrained than trained words may result from the greater effort that may be required to recall information (e.g., attempting to explicitly retrieve possible rules and the contexts they should be applied in) for novel forms, as compared to retrieving already-learned words (Ullman, 2004, 2006, 2016).

The behavioral results concerning the role of phonological cues suggest that producing new inflected forms relies to some extent on such cues. The finding that participants rely on phonological regularities in inflecting untrained words is consistent with findings from a larger behavioral study that used several variations of the same artificial language (Nevat et al., under review). In that study, as in the present one, phonological cues contributed to participants' responses to untrained items. The current results are also consistent with natural language studies that provide evidence for a reliance on phonological similarity for affixal inflection in children (Laaha, 2011; Marchman, 1997; Hartshorne and Ullman, 2006) and adults (e.g., Albright and Hayes, 2003), though it is debated whether such processes are compositional or analogic (Marchman, 1997; Hartshorne and Ullman, 2006; Albright and Hayes, 2003). The results of the present study support a role for phonological cues in compositional processes,

since the same regions implicated in this study in inflecting novel forms based on phonological cues are also involved more broadly in compositional grammatical processes, including syntax (see above).

As we have seen, proportions of “optimal” responses in the first session correlated positively with activation in pre-SMA in the same session. Activation in pre-SMA has been reported during the processing of learned complex finger-tapping sequences (Lehéricy et al., 2006). Pre-SMA is also implicated in the selection of “action sets”, i.e., responding to cues indicating changes in task requirements (Rushworth et al., 2004), such as a change in the mapping of stimuli to responses (Dove et al., 2000). It has been suggested that the involvement of pre-SMA in the processing of complex sequences and in the selection among conflicting responses are both related to the complexity of condition-action associations (Nachev et al., 2008). The fact that activity in pre-SMA was stronger among participants who relied more on rime cues in the present experiment is consistent with this view, as the selection of the optimal responses in the current study involved complex condition-action associations (i.e., between the stem and selection of the optimal affix). Moreover, the correlation suggests that the compositional processes that appear to underlie the production of untrained words, and that seem to rely on procedural memory, depend at least partly on phonological regularities.

Activation in the left pars triangularis in the first session correlated negatively with proportions of optimal responses in the third session. The pars triangularis may play roles in the recall of information from declarative memory (see above), as well as encoding such information (Chein and Fiez, 2001), and the maintenance of information in phonological working memory (Chein and Fiez, 2001; Nixon et al., 2004). It is thus possible that the activation in this region in the present experiment is related to the one or more of these functions: e.g., the recall of information from declarative memory (see above), or perhaps the maintenance of presented words in working memory till a response (inflection) is formulated. The negative correlation between activation in the left pars triangularis and proportions of optimal responses suggests that a greater reliance on such processes early in learning predicts less reliance at later stages on phonologically modulated affixation which may depend on procedural memory. Indeed, this interpretation is consistent with the view that learning in declarative memory may inhibit learning in procedural memory (Poldrack and Packard, 2003; Ullman, 2016).

4.2. Affixal inflection of trained words

As also summarized at the beginning of the Discussion, the behavioral results revealed that performance on trained items improved over the course of training. Additionally, affix type frequency modulated accuracy and reaction times, with the highest accuracy found on inflected forms taking the high-frequency inflection, and the slowest response times observed on inflected forms taking the medium-frequency inflection, with no speed-accuracy tradeoffs.

The imaging results revealed activation for inflecting trained words (vs. repeating them) in medial frontal regions, left inferior frontal regions, the left and right heads of the caudate nuclei, as well as occipital cortex, but only in the first session. Indeed, activation in the left and right caudate heads was found *only* for trained items, and only in the first session, with no activation for untrained items. Additionally, activation for trained items in the first session was modulated significantly by suffix type frequency in left inferior frontal regions, and marginally significantly in medial frontal cortex and the left head of the caudate nucleus: these regions showed the least activation for words inflected with the high-frequency suffix, with no activation differences between words inflected with the medium- and low-frequency suffixes.

As we have seen, activation in the caudate nuclei (left and right caudate heads) was specific for the inflection of trained items, and was evident only in the first session, a time at which participants were at early stages of learning. This pattern of activation is consistent with learning the regularities of affixation in procedural memory, as predicted by the DP model. Consistent with this prediction, previous studies have reported activation of the caudate nuclei bilaterally during artificial grammar learning tasks (Lieberman et al., 2004; Forkstam et al., 2006) and in non-native compared to native speakers in syntactic processing (Ruschemeyer et al., 2005). Moreover, in non-linguistic procedural learning tasks the caudate nuclei have been found to be active during early stages of category learning (Poldrack et al., 2001; Seger and Cincotta, 2005) and during the implicit learning of motor and perceptual sequences (Doyon and Benali, 2005; Gheysen et al., 2011). The caudate nuclei, in particular the caudate head bilaterally, have also been implicated in specific language impairment, in which it has been suggested that procedural memory impairments lead to the grammatical deficits found in the disorder (Ullman and Pierpont, 2005; Ullman et al., in preparation). Interestingly, previous studies have reported both anatomical (Lehéricy et al., 2004) and functional (Kim et al., 2010) connections between the caudate nuclei and pre-SMA, which we have argued above is involved in procedural memory-based compositional processes that depend on phonological cues. Overall, the activation observed in the caudate nucleus thus seems to support the learning of (phonologically cued) affix regularities in procedural memory, which are processed in frontal regions, especially pre-SMA.

The behavioral results showing an effect of suffix type frequency are consistent with our independent behavioral study of similar artificial languages conducted on a larger sample of participants (Nevat et al., under review). In that study learners showed the highest accuracy for those trained words that were inflected with the high-frequency suffix, and the lowest accuracy and slower responses for trained words inflected with the medium-frequency suffix. In the present study, the high accuracy of trained words with high-frequency affixes is also consistent with the lowest levels of activation (indeed, virtually no activation in most cases) on these words in the regions listed above (left inferior frontal, medial frontal, left caudate head). Note that as discussed above, the finding that the left inferior and medial frontal regions showed more activation for inflecting untrained than trained words suggests that they may underlie the procedural memory-based *composition* of these forms. Additionally, we have seen that the activation of the caudate nuclei on trained words suggests the involvement of these nuclei in *learning* affixal regularities in procedural memory. Thus, the weak activation in the first session of trained items

with the high-frequency suffix in these regions (left inferior frontal, medial frontal, left caudate head) may suggest that their inflection relies less on procedural memory (in learning and composition) than items inflected with medium and low frequency suffixes, perhaps relying on storage instead. More generally, the weaker activation of trained than untrained items in these regions suggests a lower involvement of composition and increased storage for trained as compared to untrained items. Indeed, the absence of any differences in activation in these regions (or other regions) between trained inflected forms and repeated forms in the third session suggests that the two types of forms are processed similarly at this point, consistent with the lack of composition and likely storage of these inflected forms. Thus, not much over a dozen presentations of each inflected form (across the instruction and training blocks) may be sufficient for them to be stored to the point where their retrieval is more efficient than composing them on-line (at least with the small set of stimuli used in the current study). Interestingly, previous studies indicate that high token frequency (high frequency of the inflected form) may contribute to the storage of affixed inflected forms (e.g., Alegre and Gordon, 1999; Dye et al., 2013; Prado and Ullman, 2009). The current results suggest that even high *affix type frequency* (high frequency of the affix) seems to increase the likelihood that inflected forms are stored, though the mechanisms of such an effect remain unclear.

An intriguing set of findings related to effects of suffix type frequency were the performance and brain activation patterns of trained words receiving the medium-frequency suffix. Performance on these words was slower, and accuracy was no higher, than on words receiving the low-frequency suffix. Moreover, brain regions showing effects of suffix frequency (left inferior frontal, medial frontal, left caudate head) showed no differences in activation between words receiving the medium- and low-frequency suffixes, despite the higher frequency of the medium-frequency suffix (Fig. 9). Also, more robust activation differences were found between items taking high- and medium-frequency affixes than between those taking high- and low-frequency affixes (Fig. 9). Overall, these results could be explained by the somewhat lower predictability of the medium-frequency suffix as compared to the (average) predictability of the low-frequency suffixes. Thus, the results suggest the possibility that learning affixal inflection (i.e., as examined in the trained items) may be modulated not only by suffix frequency, but also by the predictability of the suffix given phonological cues. Nevertheless, this possibility should be treated with caution, especially because the differences in predictability between the low- and medium-suffixes were quite small.

4.3. Limitations

Despite the advantages of using artificial languages in providing control over the participants' input and in their potential achievement of high levels of performance, such paradigms also have limitations regarding their ecological validity (see Introduction). In particular, in the current study the relatively small number of words in the language might have affected the likelihood of learning the phonological regularities, and therefore could have affected the learning mechanisms involved. Furthermore, participants' ability to learn the inflections in the current study may have been affected by the pattern of inflections in their native language. There is now considerable behavioral and neural evidence suggesting that cross-language similarity can improve the morpho-syntactic processing of a second language (Tolentino and Tokowicz, 2011; Sabourin and Stowe, 2008; Gillon Dowens et al., 2010; Renner, 2014). For the Hebrew native speakers in the present study, learning the association of a phonological cue in word-end position with a plural suffix may have been facilitated by the similar properties of their first language (Berent et al., 1999; Ravid et al., 2008). However, unlike the artificial language used here, Hebrew has only two plural suffixes, and inflections are generally determined by the grammatical gender of the singular form (though gender is associated

with phonological cues- Berent et al., 1999; Ravid et al., 2008). Thus, as in any study of learning an additional language, differences and similarities between Hebrew and the artificial language could have affected learning.

5. Conclusion

In a well-controlled artificial language paradigm, the current study revealed neural substrates involved in the learning and generalization of affixal inflectional morphology, and the roles of affix type frequency and affix phonological predictability in the modulation of these substrates.

The findings showed that medial and inferior frontal regions and the caudate nuclei played important roles in these processes. In particular, the pattern of findings suggests the importance of procedural memory and its brain structures: the learning of affixal regularities appears to involve the head of the caudate nucleus, bilaterally, whereas, after learning, the compositional processing of these regularities (as observed in the generalization of learned inflections to untrained items) seems to depend on frontal regions involved in procedural memory, including the pars opercularis, the precentral gyrus, and pre-SMA. At least pre-SMA seems to underlie aspects of composition that depend on phonological cues. Procedural memory may play less of a role in the inflection of trained words, in particular those with high-frequency suffixes, which instead might depend more on declarative memory. Indeed, learning trained words in declarative memory may inhibit the phonologically-modulated composition of inflected forms by procedural memory, as evidenced by negative correlations between activation in pars triangularis (involved in recall in declarative memory) and composition modulated by phonological cues.

Overall, the findings are consistent with the predictions of the declarative/procedural model, and further specify the model. Specifically, the results reveal roles for both affix type frequency and affix phonological predictability in modulating the involvement of the memory systems in learning and generalization. Noteworthy, however, is the fact that we did not find activation in medial temporal structures, whose involvement is expected by the model in learning new information in declarative memory. The observed pattern may be explained by the fact that participants were scanned for the first time only at the end of the first day, after a fair amount of training, since previous findings suggest that activation in these structures may occur only in very early stages of learning (Breitenstein et al., 2005; Davis and Gaskell, 2009).

In conclusion, the present study elucidates the neurocognitive bases of the acquisition and generalization of affixal inflectional morphology. Given the role of affixal morphology as a model for grammar learning, the study may potentially inform our understanding of grammar learning in language more generally.

Acknowledgments

This study was funded by BSF grant no. 2007077 to Tali Bitan and Michael Ullman.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.neuropsychologia.2016.08.026>.

References

Abutalebi, J., 2008. Neural aspects of second language representation and language control. *Acta Psychol.* 128 (3), 466–478. <http://dx.doi.org/10.1016/j.actpsy.2008.03.014>, PMID:18479667.

Albright, A., Hayes, B., 2003. Rules vs. analogy in English past tenses: a computational/

experimental study. *Cognition* 90 (2), 119–161. [http://dx.doi.org/10.1016/S0010-0277\(03\)00146-X](http://dx.doi.org/10.1016/S0010-0277(03)00146-X).

Alegre, M., Gordon, P., 1999. Frequency effects and representational status of regular inflections. *J. Mem. Lang.* 40, 41–61. <http://dx.doi.org/10.1006/jmla.1998.2607>.

Ashby, F.G., Ennis, J.M., Spiering, B.J., 2007. A neurobiological theory of automaticity in perceptual categorization. *Psychol. Rev.* 114 (3), 632–656.

Berent, I., Pinker, S., Shimron, J., 1999. Default nominal inflection in Hebrew: evidence for mental variables. *Cognition* 72 (1), 1–44. [http://dx.doi.org/10.1016/S0010-0277\(99\)00027-X](http://dx.doi.org/10.1016/S0010-0277(99)00027-X).

Beretta, A., Campbell, C., Carr, T., Huang, J., Schmitt, L., Christianson, K., Cao, Y., 2003. An ER-fMRI investigation of morphological inflection in German reveals that the brain makes a distinction between regular and irregular forms. *Brain Lang.* 85 (1), 67–92. [http://dx.doi.org/10.1016/S0093-934X\(02\)00560-6](http://dx.doi.org/10.1016/S0093-934X(02)00560-6).

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.F., Sommer, J., Wolbers, T., Knecht, S., 2005. Hippocampus activity differentiates good from poor learners of a novel lexicon. *Neuroimage* 25 (3), 958–968. <http://dx.doi.org/10.1016/j.neuroimage.2004.12.019>.

Brett, M., Anton, J., Valabregue, R., Poline, J., 2002. Region of interest analysis using an SPM toolbox Paper presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan.

Bybee, J., Newman, J., 1995. Are stem changes as natural as affixes? *Linguistics* 33, 633–654. <http://dx.doi.org/10.1515/ling.1995.33.4.633>.

Chein, J.M., Fiez, J.A., 2001. Dissociation of verbal working memory system components using a delayed serial recall task. *Cereb. Cortex* 11 (11), 1003–1014. <http://dx.doi.org/10.1093/cercor/11.11.1003>, PMID:11590110.

Clahsen, H., Felser, C., 2006. Grammatical processing in language learners. *Appl. Psycholinguist.* 27, 3–42. <http://dx.doi.org/10.1017/S0142716406060024>.

Croft, W., 2007. Construction grammar. In: Geeraerts, D., Cuyckens, H. (Eds.), *The Oxford Handbook of Cognitive Linguistics*. Oxford University Press, New York, 463–508. <http://dx.doi.org/10.1093/oxfordhb/9780199738632.001.0001>.

Dabrowska, E., Szczerbiński, M., 2006. Polish children's productivity with case marking: the role of regularity, type frequency, and phonological diversity. *J. Child Lang.* 33 (3), 559–597. <http://dx.doi.org/10.1017/S0305000906007471>, PMID:17017279.

Davis, M.H., Gaskell, M.G., 2009. A complementary systems account of word learning: neural and behavioural evidence. *Philos. Trans. R. Soc. B-Biol. Sci.* 364 (1536), 3773–3800. <http://dx.doi.org/10.1098/rstb.2009.0111>.

Davis, M.H., Di Betta, A.M., Macdonald, M.J.E., Gaskell, M.G., 2009. Learning and consolidation of novel spoken words. *J. Cognit. Neurosci.* 21 (4), 803–820.

Desai, R., Conant, L., Waldron, E., Binder, J., 2006. fMRI of past tense processing: the effects of phonological complexity and task difficulty. *J. Cognit. Neurosci.* 18 (2), 278–297. <http://dx.doi.org/10.1162/jocn.2006.18.2.278>, PMID:16494687

Dhond, R., Marinkovic, K., Dale, A., Witzel, T., Halgren, E., 2003. Spatiotemporal maps of past-tense verb inflection. *Neuroimage* 19 (1), 91–100. [http://dx.doi.org/10.1016/S1053-8119\(03\)00047-8](http://dx.doi.org/10.1016/S1053-8119(03)00047-8).

Dove, A., Pollmann, S., Schubert, T., Wiggins, C.J., von Cramon, D.Y., 2000. Prefrontal cortex activation in task switching: an event-related fMRI study. *Cognit. Brain Res.* 9 (1), 103–109. [http://dx.doi.org/10.1016/S0926-6410\(99\)00029-4](http://dx.doi.org/10.1016/S0926-6410(99)00029-4).

Doyon, J., Benali, H., 2005. Reorganization and plasticity in the adult brain during learning of motor skills. *Curr. Opin. Neurobiol.* 15 (2), 161–167. <http://dx.doi.org/10.1016/j.conb.2005.03.004>, PMID:15831397.

Dye, C.D., Walenski, M., Prado, E.L., Mostofsky, S., Ullman, M.T., 2013. Children's computation of complex linguistic forms: a study of frequency and imageability effects. *PLoS ONE* 8 (9). <http://dx.doi.org/10.1371/journal.pone.0074683>, PMID:24040318 PMID:PMC3767641.

Eichenbaum, H., 2003. Memory systems. In: Gallagher, M., Nelson, R. (Eds.), *Handbook of Psychology* 3. John Wiley & Sons, Hoboken, 543–559. <http://dx.doi.org/10.1002/0471264385.wei0319>.

Ellis, N., Schmidt, R., 1998. Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Lang. Cognit. Process.* 13 (2–3), 307–336. <http://dx.doi.org/10.1080/016909698386546>.

Ellis, N.C., Schmidt, R., 1997. Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. *Stud. Second Lang. Acquisit.* 02, 145–171.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., Wong, P.C.M., 2016. The relationship between artificial and second language learning. *Cognit. Sci.* 40, 822–847. <http://dx.doi.org/10.1111/cogs.12257>.

Field, A., 2005. *Statistics Using SPSS 2nd ed.*. SAGE Publications Ltd., London.

Forkstam, C., Hagoort, P., Fernandez, G., Ingvar, M., Petersson, K., 2006. Neural correlates of artificial syntactic structure classification. *Neuroimage* 32 (2), 956–967. <http://dx.doi.org/10.1016/j.neuroimage.2006.03.057>, PMID:16757182.

Friederici, A.D., 2011. The brain basis of language processing: from structure to function. *Physiol. Rev.* 91 (4), 1357–1392. <http://dx.doi.org/10.1152/physrev.00006.2011>.

Gabrieli, J., 1998. Cognitive neuroscience of human memory. *Annu. Rev. Psychol.* 49, 87–115. <http://dx.doi.org/10.1146/annurev.psych.49.1.87>, PMID:9496622.

Gheysen, F., Van Opstal, F., Roggeman, C., Van Waelvelde, H., Fias, W., 2011. The neural basis of implicit perceptual sequence learning. *Front. Hum. Neurosci.* 5. <http://dx.doi.org/10.3389/fnhum.2011.00137>, PMID:22087090 PMID:PMC3213531.

Gillon Dowens, M., Vergara, M., Barber, H.A., Carreiras, M., 2010. Morphosyntactic Processing in Late Second-Language Learners. *J. Cognit. Neurosci.* 22 (8), 1870–1887. <http://dx.doi.org/10.1162/jocn.2009.21304>.

Hartshorne, J.K., Ullman, M.T., 2006. Why girls say 'helded' more than boys. *Dev. Sci.* 9 (1), 21–32. <http://dx.doi.org/10.1111/j.1467-7687.2005.00459.x>, PMID:16445392.

Heim, S., Alter, K., Friederici, A.D., 2005. A dual-route account for access to grammatical gender: evidence from functional MRI. *Anat. Embryol.* 210 (5–6), 473–483. <http://dx.doi.org/10.1007/s00429-005-0032-6>, PMID:16180020.

- Henson, R., Buchel, C., Josephs, O., Friston, K., 1999. The slice timing problem in event-related fMRI. Poster presented at the 5th International Conference on Functional Mapping of the Human Brain, Dusseldorf.
- Hernandez, A., Li, P., MacWhinney, B., 2005. The emergence of competing modules in bilingualism. *Trends Cognit. Sci.* 9 (5), 220–225. <http://dx.doi.org/10.1016/j.tics.2005.03.003>, PMID:15866148 PMCID:PMC4107303.
- Karuza, E.A., Newport, E.L., Aslin, R.N., Starling, S.J., Tivarus, M.E., Bavelier, D., 2013. The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain Lang.* 127 (1), 46–54. <http://dx.doi.org/10.1016/j.bandl.2012.11.007>.
- Kempe, V., Brooks, P.J., 2008. Second language learning of complex inflectional systems. *Language Learning* 58 (4), 703–746.
- Kim, J.H., Lee, J.M., Jo, H.J., Kim, S.H., Lee, J.H., Kim, S.T., Seo, S.W., Cox, R.W., Na, D.L., Kim, S.I., Saad, Z.S., 2010. Defining functional SMA and pre-SMA subregions in human MFC using resting state fMRI: Functional connectivity-based parcellation method. *NeuroImage* 49 (3), 2375–2386. <http://dx.doi.org/10.1016/j.neuroimage.2009.10.016>, PMID:19837176 PMCID:PMC2819173.
- Köpcke, K., 1998. The acquisition of plural marking in English and German revisited: Schemata versus rules. *J. Child Lang.* 25 (2), 27. <http://dx.doi.org/10.1017/S0305000998003407>.
- Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.J., 2009. Circular analysis in systems neuroscience: The dangers of double dipping. *Nat. Neurosci.* 12 (5), 535–540. <http://dx.doi.org/10.1038/nn.2303>.
- Laaha, S., 2011. Sonority, gender and the impact of suffix predictability on the acquisition of German noun plurals. *Lang., Interact. Acquis.* 2 (1), 82–100. <http://dx.doi.org/10.1075/lia.2.1.04laa>.
- Leach, L., Samuel, A.G., 2007. Lexical configuration and lexical engagement: when adults learn new words. *Cognit. Psychol.* 55, 306–353.
- Lehéricy, S., Ducros, M., Krainik, A., Francois, C., Van de Moortele, P.F., Ugurbil, K., Kim, D.S., 2004. 3-D diffusion tensor axonal tracking shows distinct SMA and Pre-SMA projections to the human striatum. *Cereb. Cortex* 14 (12), 1302–1309. <http://dx.doi.org/10.1093/cercor/bhh091>, PMID:15166103.
- Lehéricy, S., Bardinet, E., Tremblay, L., Van de Moortele, P.F., Pochon, J.P., Dormont, D., Kim, D.S., Yelnik, J., Ugurbil, K., 2006. Motor control in basal ganglia circuits using fMRI and brain atlas approaches. *Cereb. Cortex* 16, 149–161. <http://dx.doi.org/10.1093/cercor/bhi089>, PMID:15858164.
- Lieberman, M., Chang, G., Chiao, J., Bookheimer, S., Knowlton, B., 2004. An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *J. Cognit. Neurosci.* 16 (3), 427–438. <http://dx.doi.org/10.1162/089929004322926764>, PMID:15072678.
- Marchman, V.A., 1997. Children's productivity in the English past tense: The role of frequency, phonology and neighborhood structure. *Cognit. Sci.* 21 (3), 283–304. http://dx.doi.org/10.1207/s15516709cog2103_2.
- Marcus, G., Brinkmann, U., Clahsen, H., Wiese, R., Pinker, S., 1995. German inflection: the exception that proves the rule. *Cognit. Psychol.* 29 (3), 189–256. <http://dx.doi.org/10.1006/cogp.1995.1015>, PMID:8556846.
- Maslen, R., Theakston, A., Lieven, E., Tomasello, M., 2004. A dense corpus study of past tense and plural overregularization in English. *J. Speech Lang. Hear. Res.* 47 (6), 1319–1333. [http://dx.doi.org/10.1044/1092-4388\(2004\)099](http://dx.doi.org/10.1044/1092-4388(2004)099).
- Merkx, M., Rastle, K., Davis, M.H., 2011. The acquisition of morphological knowledge investigated through artificial language learning. *Q. J. Exp. Psychol.* 64 (6), 1200–1220. <http://dx.doi.org/10.1080/17470218.2010.538211>.
- Morgan-Short, K., Steinhauer, K., Sanz, C., Ullman, M.T., 2012a. Explicit and implicit second language differentially affect achievement of native-like brain activation patterns. *J. Cognit. Neurosci.* 24 (4), 933–947. <http://dx.doi.org/10.1162/jocn-a-00119>.
- Morgan-Short, K., Finger, I., Grey, S., Ullman, M.T., 2012b. Second language processing shows increased native-like neural responses after months of no exposure. *PLoS One* 7 (3). <http://dx.doi.org/10.1371/journal.pone.0032974>.
- Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., 2003. Broca's area and the language instinct. *Nat. Neurosci.* 6 (7), 774–781. <http://dx.doi.org/10.1038/nn1077>, PMID:12819784.
- Nachev, P., Kennard, C., Husain, M., 2008. Functional role of the supplementary and pre-supplementary motor areas. *Nat. Rev. Neurosci.* 9 (11), 856–869. <http://dx.doi.org/10.1038/nrn2478>, PMID:18843271.
- Nevat, M., Khateb, A., Prior, A., 2014. When first language is not first: an functional magnetic resonance imaging investigation of the neural basis of diglossia in Arabic. *Eur. J. Neurosci.* 40 (9), 3387–3395. <http://dx.doi.org/10.1111/ejn.12673>, PMID:25040779.
- Nevat, M., Ullman, M.T., Eviatar, Z., Bitan, T., 2016. The role of distributional factors in learning and generalizing affixal inflection: An artificial language study (under review).
- Newman, A.J., Pancheva, R., Ozawa, K., Neville, H.J., Ullman, M.T., 2001. An event-related fMRI study of syntactic and semantic violations. *J. Psycholinguist.* Res. 30 (3), 339–364. <http://dx.doi.org/10.1023/A:1010499119393>, PMID:11523278.
- Newman-Norlund, R., Frey, S., Petitto, L., Grafton, S., 2006. Anatomical substrates of visual and auditory miniature second-language learning. *J. Cognit. Neurosci.* 18 (12), 1984–1997. <http://dx.doi.org/10.1162/jocn.2006.18.12.1984>, PMID:17129186 PMCID:PMC1868893.
- Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., Passingham, R., 2004. The inferior frontal gyrus and phonological processing: an investigation using rTMS. *J. Cognit. Neurosci.* 16 (2), 289–300. <http://dx.doi.org/10.1162/089892904322984571>, PMID:15068598.
- Oh, T.M., Tan, K.L., Ng, P., Berne, Y.I., Graham, S., 2011. The past tense debate: Is phonological complexity the key to the puzzle? *Neuroimage* 57 (1), 271–280. <http://dx.doi.org/10.1016/j.neuroimage.2011.04.008>, PMID:21511040.
- Opitz, B., Friederici, A.D., 2003. Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *Neuroimage* 19 (4), 1730–1737. [http://dx.doi.org/10.1016/S1053-8119\(03\)00170-8](http://dx.doi.org/10.1016/S1053-8119(03)00170-8).
- Paradis, M., 1994. Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In: Ellis, N.C. (Ed.), *Implicit and explicit learning of languages*. Academic Press, London, UK, 393–419.
- Pinker, S., 1991. Rules of language. *Science* 253 (5019), 530–535. <http://dx.doi.org/10.1126/science.1857983>, PMID:1857983.
- Pinker, S., Ullman, M., 2002. The past and future of the past tense. *Trends Cognit. Sci.* 6 (11), 456–463. [http://dx.doi.org/10.1016/S1364-6613\(02\)01990-3](http://dx.doi.org/10.1016/S1364-6613(02)01990-3).
- Pliatsikas, C., Johnstone, T., Marinis, T., 2014. fMRI evidence for the involvement of the procedural memory system in morphological processing of a second language. *PLoS One* 9 (5), e97298. <http://dx.doi.org/10.1371/journal.pone.0097298>.
- Poldrack, R., Clark, J., Paré-Blagoev, E., Shohamy, D., Creso Moyano, J., Myers, C., Gluck, M., 2001. Interactive memory systems in the human brain. *Nature* 414 (6863), 546–550. <http://dx.doi.org/10.1038/35107080>, PMID:11734855.
- Poldrack, R.A., Packard, M.G., 2003. Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia* 41 (3), 245–251.
- Prado, E.L., Ullman, M.T., 2009. Can imageability help us draw the line between storage and composition? *J. Exp. Psychol. Learn., Mem. Cogn.* 35 (4), 849–866. <http://dx.doi.org/10.1037/a0015286>, PMID:19586257.
- Ravid, D., Dressler, W.U., Nir-Sagiv, B., Korecky-Kröll, K., Souman, A., Rehfeldt, K., Laaha, S., Bertl, J., Basboell, H., Gillis, S., 2008. Core morphology in child directed speech: crosslinguistic corpus analyses of noun plurals. In: Behrens, H. (Ed.), *Corpora in language acquisition research: finding structure in data*. Benjamins, Amsterdam, 25–60.
- Renner, A., 2014. L1 transfer effects in L2 grammatical gender processing of late bilinguals (doctoral dissertation). Retrieved from Humboldt University's Publication Server. (<http://edoc.hu-berlin.de/dissertationen/renner-anna-2014-04-14/PDF/renner.pdf>).
- Ruschmeyer, S.A., Fiebach, C.J., Kempe, V., Friederici, A.D., 2005. Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Hum. Brain Mapp.* 25 (2), 266–286. <http://dx.doi.org/10.1002/hbm.20098>, PMID:15849713.
- Rushworth, M.F., Walton, M.E., Kennerly, S.W., Bannerman, D.M., 2004. Action sets and decisions in the medial frontal cortex. *Trends Cognit. Sci.* 8 (9), 410–417. <http://dx.doi.org/10.1016/j.tics.2004.07.009>, PMID:15350242.
- Sabourin, L., Stowe, L.A., 2008. Second language processing: when are first and second languages processed similarly? *Second Lang. Res.* 24 (3), 397–430. <http://dx.doi.org/10.1177/02676583080809186>.
- Sahin, N., Pinker, S., Halgren, E., 2006. Abstract grammatical processing of nouns and verbs in Broca's area: evidence from fMRI. *Cortex* 42 (4), 540–562. [http://dx.doi.org/10.1016/S0010-9452\(08\)70394-0](http://dx.doi.org/10.1016/S0010-9452(08)70394-0).
- Seger, C.A., Cincotta, C.M., 2005. The roles of the caudate nucleus in human classification learning. *J. Neurosci.* 25 (11), 2941–2951. <http://dx.doi.org/10.1523/JNEUROSCI.3401-04.2005>, PMID:15722354.
- Serences, J.T., 2004. A comparison of methods for characterizing the event-related BOLD timeseries in rapid fMRI. *NeuroImage* 21, 1690–1700. <http://dx.doi.org/10.1016/j.neuroimage.2003.12.021>, PMID:15050591.
- Squire, L., 2004. Memory systems of the brain: A brief history and current perspective. *Neurobiol. Learn. Mem.*, 82, 171–177. <http://dx.doi.org/10.1016/j.nlm.2004.06.005>, PMID:15464402.
- Tagarelli, K., Grey, S., Turkeltaub, P.E., Ullman, M.T., 2016. The functional neuroanatomy of adult second language: an activation likelihood estimation meta-analysis (in preparation).
- Tamminen, J., Davis, M.H., Merkx, M., Rastle, K., 2012. The role of memory consolidation in generalisation of new linguistic information. *Cognition* 125 (1), 107–112. <http://dx.doi.org/10.1016/j.cognition.2012.06.014>.
- Tatsuno, Y., Sakai, K.L., 2005. Language-related activations in the left prefrontal regions are differentially modulated by age, proficiency, and task demands. *J. Neurosci.* 25 (7), 1637–1644. <http://dx.doi.org/10.1523/JNEUROSCI.3978-04.2005>, PMID:15716399.
- Tolentino, L.C., Tokowicz, N., 2011. Across languages, space, and time: a review of the role of cross-language similarity in L2 (Morpho) syntactic processing as revealed by fMRI and ERP methods. *Stud. Second Lang. Acquis.* 33 (1), 91–125. <http://dx.doi.org/10.1017/s0272263110000549>.
- Tyler, L., Stamatakis, E., Post, B., Randall, B., Marslen-Wilson, W., 2005. Temporal and frontal systems in speech comprehension: an fMRI study of past tense processing. *Neuropsychologia* 43 (13), 1963–1974. <http://dx.doi.org/10.1016/j.neuropsychologia.2005.03.008>, PMID:16168736.
- Ullman, M., 2001a. The declarative/procedural model of lexicon and grammar. *J. Psycholinguist. Res.* 30 (1), 37–69. <http://dx.doi.org/10.1023/A:1005204207369>, PMID:11291183.
- Ullman, M., 2004. Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92 (1–2), 231–270. <http://dx.doi.org/10.1016/j.cognition.2003.10.008>, PMID:15037131.
- Ullman, M., 2006. Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex* 42 (4), 480–485. [http://dx.doi.org/10.1016/S0010-9452\(08\)70382-4](http://dx.doi.org/10.1016/S0010-9452(08)70382-4).
- Ullman, M.T., 2001b. The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Biling. Lang. Cogn.* 4 (1), 105–122. <http://dx.doi.org/10.1017/s1366728901000220>.
- Ullman, M.T., 2005. A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In: Sanz, C. (Ed.), *Mind and context in adult second language acquisition: Methods, theory and practice*. Georgetown University Press, Washington DC, 141–178.
- Ullman, M.T., 2015. The declarative/procedural model: A neurobiologically motivated

- theory of first and second language. In: VanPatten, B., Williams, J. (Eds.), *Theories in Second Language Acquisition: An Introduction* 2nd ed.. Routledge, New York, 135–158.
- Ullman, M.T., 2016. The declarative/procedural model: A neurobiological model of language learning, knowledge and use. In: Hickok, G., Small, S.A. (Eds.), *The Neurobiology of Language*. Elsevier, 953–968.
- Ullman, M.T., Pierpont, E.I., 2005. Specific language impairment is not specific to language: the procedural deficit hypothesis. *Cortex* 41 (3), 399–433.
- Ullman M.T., Pullman M.T., Lovelett J.T., Pierpont E.I., Turkeltaub P.E., Brain abnormalities in Specific Language Impairment are localized to frontal regions and the caudate nucleus, 2016a (in preparation)
- Ullman, M.T., Campbell, R., McQuaid, G.A., Tagarelli, K.M., Turkeltaub, P.E., 2016b. The functional neuroanatomy of regular and irregular morphology (in preparation)
- Vannest, J., Polk, T.A., Lewis, R.L., 2005. Dual-route processing of complex words: New fMRI evidence from derivational suffixation. *Cognit. Affect. Behav. Neurosci.* 5 (1), 67–76. <http://dx.doi.org/10.3758/CABN.5.1.67>, PMID:15913009.
- Wartenburger, I., Heekeren, H., Abutalebi, J., Cappa, S.F., Villringer, A., Perani, D., 2003. Early setting of grammatical processing in the bilingual brain. *Neuron* 37, 159–170. [http://dx.doi.org/10.1016/S0896-6273\(02\)01150-9](http://dx.doi.org/10.1016/S0896-6273(02)01150-9).
- Zaccarella, E., Friederici, A., 2015. Merge in the human brain: A sub-region based functional investigation in the left pars opercularis. *Front. Psychol.*, 6. <http://dx.doi.org/10.3389/fpsyg.2015.01818>.