Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization

Atsuko Takashima [a,b,*], Iske Bakker [a,b], Janet G. van Hell [c,a], Gabriele Janzen [a,b], James M. McQueen [a,b,d]

[a] *Radboud University Nijmegen, Behavioural Science Institute, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands*
[b] *Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands*
[c] *Pennsylvania State University, Department of Psychology, University Park, PA 16802, USA*
[d] *Max Planck Institute for Psycholinguistics, Nijmegen, P.O. Box 310, 6500 AH Nijmegen, The Netherlands*

ARTICLE INFO

ABSTRACT

The complementary learning systems account of declarative memory suggests two distinct memory networks, a fast-mapping, episodic system involving the hippocampus, and a slower semantic memory system distributed across the neocortex in which new information is gradually integrated with existing representations. In this study, we investigated the extent to which these two networks are involved in the integration of novel words into the lexicon after extensive learning, and how the involvement of these networks changes after 24 h. In particular, we explored whether having richer information at encoding influences the lexicalization trajectory. We trained participants with two sets of novel words, one where exposure was only to the words' phonological forms (the form-only condition), and one where pictures of unfamiliar objects were associated with the words' phonological forms (the picture-associated condition). A behavioral measure of lexical competition (indexing lexicalization) indicated stronger competition effects for the form-only words. Imaging (fMRI) results revealed greater involvement of phonological lexical processing areas immediately after training in the form-only condition, suggesting that tight connections were formed between novel words and existing lexical entries already at encoding. Retrieval of picture-associated novel words involved the episodic/hippocampal memory system more extensively. Although lexicalization was weaker in the picture-associated condition, overall memory strength was greater when tested after a 24 hour delay, probably due to the availability of both episodic and lexical memory networks to aid retrieval. It appears that, during lexicalization of a novel word, the relative involvement of different memory networks differs according to the richness of the information about that word available at encoding.

Introduction

The ability to acquire novel words from a limited number of exposures is one of the most remarkable capacities of the human brain. For example, young children are able to associate novel word-forms with meaning extracted from context extremely rapidly, after only a single or very few encounters with a novel word ('fast mapping'; Carey and Bartlett, 1978). But is a newly-learned word-form truly a *word*, identical in nature to the representation of other words in the learner's mental lexicon? It has been suggested that while episodic memory is sufficient for fast mapping and for retrieval of the new word-form in an explicit memory test, a slower and more gradual learning mechanism is responsible for lexicalization, that is, the integration of the novel word-form into the existing lexicon (Davis and Gaskell, 2009). On this view, a true lexical representation emerges only after a consolidation period, and it differs both in neural coding and behavioral functioning from the initially-encoded episodic memory representation. However, not much is known about how this transition from episodic to lexical memory takes place at the neural level. In this study, we sought to investigate the neural correlates of novel word learning and how the integration into the existing lexicon takes place. We focused on whether having extra information (i.e., novel pictorial information) about the novel word-form at encoding would influence the lexicalization and retrieval of learned word-forms compared to learning novel word-forms alone, both behaviorally and at the neural level.

What are the characteristics of a lexical, rather than episodic representation of a novel word? Gaskell and Dumay (2003) argued that a distinguishing feature of lexical entries is their ability to influence the retrieval of other words in the lexicon. An example of this is *lexical competition*, a process in which the recognition of a spoken word is slowed down by the simultaneous activation of phonologically-related competitor words (McQueen et al., 1994). Learned novel words can be retrieved

using both episodic and lexical memory networks. Therefore, rather than testing explicit memory for a novel word, a better way of assessing its lexical status is to test whether this word is able to enter into lexical competition with existing phonologically neighboring words. To investigate the time-course of lexical integration, Gaskell and Dumay (2003) taught participants novel word forms that overlapped with an existing base word, such as *cathedruke* for *cathedral*. After an encoding phase, participants' reaction times to the existing words were tested in a lexical competition task. Immediately after training, no difference was observed between base words with versus without a novel competitor (e.g., *cathedral* when *cathedruke* had or had not been trained). In a retest one to several days later, however, participants were slower to respond to *cathedral* than to control words without a novel competitor, suggesting that the novel word entered into lexical competition with its existing neighbor if there was a period of consolidation of several hours to days (for replications, see e.g. Davis et al., 2009; Dumay and Gaskell, 2007, 2012; Lindsay and Gaskell, 2013; Tamminen and Gaskell, 2008; Tamminen et al., 2010). Since participants were able to recognize the newly-learned words immediately after training, these results support the distinction between the initial encoding process that results in an episodic, non-lexical memory, and the formation of an integrated, lexicalized representation after an additional consolidation period.

This combination of a faster episodic memory and a slower lexical memory is in line with the idea that declarative memory is composed of two complementary systems with distinct neural substrates (McClelland et al., 1995). The episodic system, relying on the medial temporal lobe (MTL) and the hippocampus in particular, is able to encode incoming information rapidly and store this information in the context of a particular episode. In this way, the MTL system is well suited for binding multiple semantically unrelated sensory inputs as being associated to a specific episode. Lexical–semantic information, on the other hand, seems to reside outside the MTL system, as semantic memories and language comprehension abilities are often spared in patients with hippocampal lesions (Alvarez and Squire, 1994; Moscovitch et al., 2006; Nadel and Moscovitch, 1997; Scoville and Milner, 1957; Squire and Zola-Morgan, 1991; Winocur and Moscovitch, 2011). Semantic memory is thus assumed to rely on a neocortically-based, distributed associative network (Martin and Chao, 2001; Patterson et al., 2007). While such a network provides a good model for existing knowledge, it is assumed that abrupt integration of new associations directly into this system would cause existing memories to be severely damaged or erased ("catastrophic interference"; McCloskey and Cohen, 1989). The complementary learning systems account avoids this problem by positing that the hippocampal system serves as a fast-learning system for new memory traces as arbitrary associations, which are subsequently slowly integrated into the neocortical network by interleaving old and new information in a more systematic and associative manner (McClelland et al., 1995).

The complementary learning systems account predicts changes in the neural representation of words as a result of consolidation: the rapid, initial learning of a novel word as supported by the MTL memory system (reflecting initial episodic memory) shifts to lexical storage in neocortical language areas with consolidation. This prediction was confirmed by a recent functional magnetic resonance imaging (fMRI) study (Davis et al., 2009), in which untrained novel words elicited the greatest hippocampal activation. But the conclusion that the novel words had been cortically integrated was not based on an increase in cortical activation for consolidated words. On the contrary, unconsolidated words elicited higher activation than consolidated words in the superior temporal gyrus (STG), an area known for prelexical phonological processing (Hickok and Poeppel, 2007) or more specifically termed the "primary site of acoustic–phonetic analysis of unmodified natural speech" (Gow, 2012). This area is known to increase in activity with a higher demand on the lexical access system, such as during processing of pseudowords over real words (Newman and Twieg, 2001). The lexical

competition task that Davis et al. used in the MRI scanner may have captured the brain activity related to demanding phonological processing rather than to the accessing of lexical representations per se. In the present study, therefore, in order to tap directly into the memory representations of the trained words, we used a memory retrieval task while we scanned participants' brain activity.

Prior studies using the Gaskell and Dumay (2003) paradigm have tended to focus on how listeners learn the phonological form of novel words. Although exposure to novel forms without meaning certainly occurs in real life, most often listeners presumably infer referents for those new words (e.g., a visual representation based on its co-occurrence with a new word). Because the memory trace for a word–object pair is semantically richer than the memory for a phonological form without meaning, one might expect lexical integration to be facilitated for object-associated words. Using a perceptual learning task to measure novel words' ability to engage with prelexical phonological representations, Leach and Samuel (2007) indeed observed that the association of word forms with contexts and pictures enhanced lexicalization.

Alternatively, however, integration of new lexical forms could be delayed for object-associated novel words relative to those learned only as phonological forms (i.e., without meaning). According to the complementary learning systems account, stronger associations between novel and existing words should facilitate the interleaving of old and new information into neocortical memory. Words without meaning are likely to be more strongly associated with existing phonological neighbors during encoding than words with new meanings. This is because in the former case the learner has only phonological information to go on (e.g., *cathedruke* sounds like *cathedral*), whereas in the latter case the learner can attend to the form–meaning relationship (e.g., the visual properties of a *cathedruke*, if it is associated with a picture). On this view, therefore, evidence of form-based lexicalization (i.e., a lexical competition effect) may emerge more rapidly for words without associated pictures than for those learned with pictures.

In the present study we aimed to elucidate the neural representation of novel words, and assess whether there is a different lexicalization trajectory in the presence versus absence of an associated visual object. We trained participants with 40 *cathedruke*-type novel spoken word forms. Half of these words were paired with pictures of novel objects (picture-associated words), and half were presented without any information other than their phonological forms (form-only words). To investigate the changes in neural representation over time, participants performed an old/new recognition task on the learned novel words intermixed with previously unheard novel words and existing words, immediately after training and again the next day, while we measured their brain activity with fMRI. Furthermore, to measure the extent of form-based lexical integration of the novel words, we administered a behavioral lexical competition task on both days.

Following Davis and Gaskell (2009), we predicted that behaviorally, lexical competition would emerge on the second day (i.e., after a period of consolidation). If richer information facilitates lexicalization (Leach and Samuel, 2007), this effect should be enhanced for picture-associated novel words compared to form-only novel words. If, in contrast, withholding meaning information leads to better interleaving of existing and novel phonological information, there should be a stronger competition effect for the form-only words than for the picture-associated words. At the neural level, in accordance with the complementary learning systems account, we expected the MTL to be involved in recognition of newly learned words on Day 1. Moreover, since picture-associated words required arbitrary visual–auditory information to be linked with each other, and given the role of the hippocampus in binding multimodal information or between-domain associations (Davachi, 2006; Mayes et al., 2007; Suzuki, 2007), we expected more involvement of the MTL system for the recognition of picture-associated words compared to form-only words, with stronger hippocampal connectivity between the auditory and visual cortices. On Day 2, we assumed overnight lexicalization to enhance the emergence of lexical entries in

the neocortical memory system, leading to a tighter link between the existing word (*cathedral*) and the newly learned word (*cathedruke*), thereby enabling stronger co-activation of base words by their novel neighbors than on Day 1. We expected this increase in lexical activation to be indexed by stronger responses in phonological lexical representational areas such as the posterior middle temporal gyrus (MTG) (Gow, 2012; Hickok and Poeppel, 2004, 2007; Prabhakaran et al., 2006), and stronger connectivity between auditory cortex and this area as a result of increased lexical integration. In summary, therefore, we explored lexicalization of newly-learned words using behavioral and imaging measures, the latter including functional connectivity measures. We focused on whether the lexicalization trajectory, as revealed by these measures, would differ depending on whether the words were learned with associated pictures or solely as phonological forms.

Methods

Participants

Twenty-five volunteers were recruited from the university volunteer panel (23 women; age range 18–28; $M = 23$; $SD = 2.8$ years; all right handed). All were native speakers of Dutch, with normal or corrected to normal vision and audition, and without any known language impairments. Each participant gave written informed consent in compliance with the local ethical committee (CMO region Arnhem–Nijmegen, The Netherlands). One female participant was excluded from further analyses due to a very low memory score (below 50% correct on any task excluding free recall) and another female participant was excluded from the fMRI analyses due to malfunctioning of the scanner during the second session.

Materials

Word-form materials consisted of digital recordings of 44 disyllabic and 76 trisyllabic word pairs, each including an existing monomorphemic Dutch noun (base word) and a pseudoword derived by substituting the final vowel(s) and consonant(s) of the base word, e.g. *kathedraal* ('cathedral') — *kathedrook*. These were thus Dutch variants of the stimuli used by Gaskell and Dumay (2003). Base words were between 5 and 9 phonemes long (mean 6.8) and had a lemma frequency between 0 and 1.69 log per million (mean 0.94; all ≥ 1 instance per million) in CELEX (Baayen et al., 1993). The uniqueness point, the phoneme at which a word diverges from all other words that initially match the input, varied between the 3rd and 6th phonemic positions and was always located before the final vowel. If exposure to the novel items generated a new lexical entry, the uniqueness point of the base word was therefore expected to shift towards the offset of the word (more precisely, to its final vowel).

The word–pseudoword pairs were divided into six lists of 20 pairs, matched on log frequency, number of phonemes and syllables, and uniqueness point. The six lists were assigned to the following conditions: form-only trained novel word, picture-associated trained novel word, untrained novel word Day 1, untrained novel word Day 2, existing word Day 1, and existing word Day 2. Assignment of lists to conditions was fully counterbalanced across participants. Six existing words based on the same criteria were included as catch trials during the training and recognition tasks.

Twenty pictures of unusual objects without a clear linguistic label were selected using the Google image database. All pictures depicted a colored object against a white background. The pairing of pictures to novel words was randomized for each participant.

Procedure

On Day 1, participants were trained through phoneme monitoring on the phonological form of 40 novel words, of which 20 were associated with pictures of unusual objects ("picture-associated words") and 20 were not ("form-only words"). They were instructed at the beginning of the session to memorize the stimuli and informed that they would be tested later. Immediately after training on Day 1, they performed a recognition test in the MRI scanner. This was followed by a lexical competition task, free recall test and picture–word source memory test outside the scanner. On departure, participants were reminded to have a regular night's sleep. Participants returned on average 25 h (range 21.5–28 h) later on Day 2, and followed the same procedure as on Day 1 but without the training phase. Training took on average 75 min, and the recognition test in the scanner took on average 25 min. The overall design is illustrated in Fig. 1.

Training: phoneme monitoring

In the training phase, participants performed a phoneme-monitoring task (as in Gaskell and Dumay, 2003) on 20 novel words from the picture-associated list, 20 novel words from the form-only list, and 6 existing words to be used as catch trials during the recognition test. Participants were told that this was a learning paradigm for new words, that they had to memorize the words, and that their memory would be tested later. For the picture-associated words they were also asked to remember (for later test) the word-picture combinations. The phoneme monitoring task consisted of 36 blocks, and the 46 words were presented once per block in randomized order. At the beginning of each block, one of six target phonemes (/k, n, m, t, p, l/, each target phoneme occurred 6 times) was presented via the headphones. Participants were instructed to press the button whenever they heard the target phoneme in a word. For words in the picture-associated condition, the picture appeared on the screen 200 ms prior to the sound onset of the word and remained on screen for 1000 ms after the offset of the word. During form-only trials a fixation cross was presented. The inter-trial interval was 700 ms. In case the participant pressed the button in target-absent trials or they missed responding within the time limit for the target-present trials, visual feedback was given to indicate inappropriate responses to increase motivation.
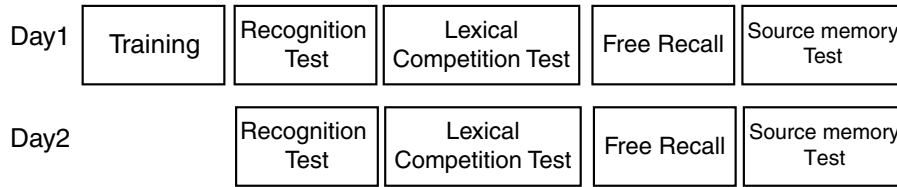
Recognition test (fMRI)

In the scanner, participants heard the 40 novel words (20 picture-associated, 20 form-only) and 6 existing words that they had been trained on, intermixed with 20 untrained novel words and 20 untrained existing words. Words were presented through in-ear headphones and the volume was adjusted for every participant such that the words were clearly audible above the scanner noise. After a jittered inter-trial interval of 2 to 6 s, a fixation cross on the screen turned from white to blue, signaling that the next word would be presented 1 s later. The participants' task was to judge each word as either old (present in the training set) or new. They indicated their response and confidence level by choosing one of six colored squares, representing a scale from 'definitely old' to 'definitely new' by moving the cursor to the right or left using button presses and then confirming with a separate button. After a jittered interval of 1 to 2 s, the participants were asked whether the word had been associated with a picture or not, and again indicated their confidence level by choosing one of six response options on a scale from 'definitely not' to 'definitely yes'. For new words, they were instructed to choose the 'definitely not' option. No feedback was given.
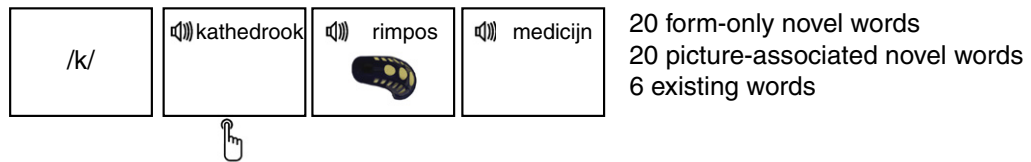
Lexical competition task: pause detection

In the lexical competition task, participants were required to detect a short silence embedded in the base words (following Gaskell and Dumay, 2003, Experiment 3). Base words of the 40 trained novel words (20 base words in the picture-associated condition, 20 base words in the form-only condition), and 20 base words from the untrained novel word condition were presented, intermixed with 60 existing word fillers not used in the training or recognition test session. In half of the words in each condition (counterbalanced across participants) a 200 ms pause was inserted just before the onset of the final
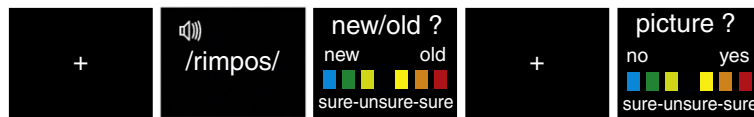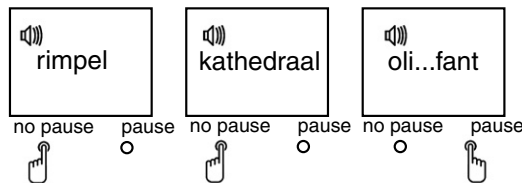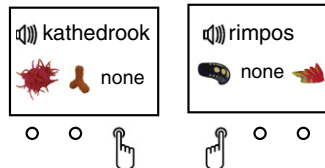
Fig. 1. Experimental design. (A) Experimental procedure. On Day 1, the phoneme monitoring task took place during training, followed by the old/new recognition test in the MRI scanner and the behavioral lexical competition and memory tasks outside the scanner. On Day 2, all tasks except phoneme monitoring were repeated. (B) Tasks. Training: participants performed a phoneme monitoring task. Recognition in the fMRI scanner: participants were instructed to make an old/new judgment on the words presented through the headphones, and whether the word accompanied a picture or not, using a three-step confidence scale for each answer option. Lexical competition task: participants were requested to detect a pause and press the relevant button. Source memory test: participants had to match the auditory words to their associated visual counterpart and press the corresponding response button.

syllable. Pauses were always placed at zero-crossings using the Praat speech editor (Boersma and Weenink, 2012). When there was no naturally occurring silence before the onset of the final syllable, such as before vowels or fricatives, a 5 ms period of the waveform before and/or after the pause was faded to ensure that no clicks were audible. For each word, participants were instructed to indicate whether there was a pause or not by pressing a pause-present button when they heard a pause and a pause-absent button when they did not detect a pause, as fast and as accurately as possible. A fixation cross remained on the screen during the task and turned green after a correct response and red after an incorrect response. 'Too late!' feedback was given when no response was made within 1000 ms after word offset. The inter-trial interval was 1000 ms.

Free recall

Participants were instructed to recall verbally as many items as they could remember of those presented in the training session, within 3 min.

Source memory test: picture–word association

All 40 trained novel words were presented via the headphones with three response options on the screen; two pictures, and a 'no picture' option. For novel words from the picture-associated condition, one of the pictures was the associated picture, and the second picture had been associated to another novel word during the training session. For form-only novel words, both pictures were associated to one of the picture-associated novel words. Participants were instructed to indicate which option was most appropriate by pressing the corresponding button. No time limit was imposed. Following each response, participants indicated their confidence level on a three-point scale. No feedback was given.

FMRI acquisition

FMRI data were recorded in a 1.5 T MR scanner (Avanto, Siemens Healthcare, Erlangen, Germany) using a 32-channel head coil. For functional images, we used a T2*-weighted gradient multi-echo planar

imaging sequence with the following parameters: repetition time (TR): 2.14 s, echo time: TE1 9.4 ms, TE2 21 ms, TE3 33 ms, TE4 44 ms, TE5 56 ms, 34 slices, ascending slice order, 3.0 mm slice thickness, 0.5 mm slice gap, matrix size: 64 × 64, field of view (FOV): 212 × 212 mm, flip angle: 90°, and voxel size: 3.0 × 3.0 × 3.5 mm. The multiple echo sequence allows less distortions and drop-outs (Poser et al., 2006). Since one of the main areas of interest, the medial temporal lobe, is susceptible to these artifacts, we opted for this sequence. Slices were angulated in an oblique axial manner to reach whole-brain coverage (except for a part of the parietal cortex). Additionally, T1-weighted anatomical scans at 1 mm isotropic resolution were acquired with TR 2250 ms, TE 2.95 ms, flip angle 15°, and FOV 256 × 256 × 176 mm.

FMRI analysis

The multi-echo sequence acquired five echoes at every time point (i.e. per TR). To allow the use of standard fMRI preprocessing tools in SPM8 (www.fil.ion.ucl.ac.uk), these five echoes were combined using the procedure described by Poser et al. (2006). Thirty volumes acquired prior to the recognition task were used to calculate the optimal weighting of the five echoes to be combined to one value per time point for every voxel, and this weighting matrix was applied to the functional scans. Image pre-processing and statistical analysis was performed using SPM8. The first five volumes of each participant's functional scan were discarded to allow for T1 equilibration. The subject-mean image of the functional run after realignment was co-registered with the corresponding structural MRI using mutual information optimization. The functional images were subsequently slice-time corrected to the first slice, using SPM8's Fourier pane shift interpolation. Structural scans were segmented and both functional and structural scans were spatially normalized and transformed into the common Montreal Neurological Institute space (resampled at voxel size 2 × 2 × 2 mm), as defined by the SPM8 T1.nii template, as well as spatially filtered by convolving the functional images with an isotropic 3D Gaussian kernel (8 mm full width at half maximum).

The fMRI data were analyzed statistically using a general linear model (GLM) and statistical parametric mapping. Six explanatory variables were included in the model for each session: form-only hits (very sure and sure), picture-associated hits (very sure and sure), untrained novel correct rejection (very sure and sure), existing words correct (very sure and sure), null events, and trials of no interest comprising of all incorrect trials, unsure correct trials responded with the least confidence level (unsure old and unsure new), and all filler trials (the six existing words that served as catch trials). These explanatory variables were temporally convolved with the canonical Hemodynamic Response Function (HRF) provided by SPM8. Each event was time-locked to the onset of the word. The design matrix included the six head motion regressors (three translations, three rotations). A high pass filter was implemented using a cut-off period of 128 s to remove low-frequency effects from the time series. For statistical analysis, relevant contrast parameter images were generated for each participant and then subjected to a second-level analysis (Penny et al., 2003), treating subjects as a random variable. To obtain ROIs for the connectivity analysis, and to perform ANOVAs on the second-level probing the changes over time for form-only and picture-associated hits, the unmodeled period (intertrial interval period) served as a common baseline to create a contrast for each of the conditions. Memory strength might decay as a function of time, and this might influence the strength of the brain response especially on Day 2. Trials with higher confidence might reflect more retrieval success than low confidence trials. If word recognition in the picture associated condition also accompanied associated picture retrieval, we might expect brain activation related to retrieval of the visual information. For these reasons, we tested two extra models on the single subject level, including a parametric modulation regressor for each of the models, one by the confidence level to the word memory response of the correctly recognized trials, and the other by the

confidence response to the picture memory question during the retrieval test in the scanner.

To probe changes in functional connectivity associated with consolidation, we performed a Psycho–Physiological Interaction (PPI) analysis (Friston et al., 1997) embedded in SPM8. In PPI analysis, functional connectivity is assessed by searching for brain regions that co-activate with a seed region (the physiological factor), and in which this co-activation is significantly different in one condition compared to another (the psychological factor). Because our stimuli were presented aurally, we selected bilateral STG areas that were activated during auditory perception of the stimuli. To define the seed region, we first defined a peak voxel in the bilateral STG on the group level that was active during all trials, and selected all significant voxels within a sphere of 10 mm centered at the peak voxel. To obtain the participants' specific fluctuation of BOLD responses to auditory input, we inclusively masked this area with subject-specific active voxels while they were attending to the trained word condition, and took those overlapping voxels with those found on the group level as the seed region of interest (ROI). In the PPI analysis, we considered areas that co-activated with the seed ROI more strongly for form-only hits than picture-associated hits and vice versa. A GLM was constructed at the single-subject level for each of the sessions separately, using three regressors: (i) the deconvolved signal from the seed ROI (Gitelman et al., 2003), (ii) trial onset for form-only hits (positive) and picture-associated hits (negative) convolved with HRF and (iii) the interaction term between the first and the second regressors, together with six condition regressors and the six movement-related regressors that were used in the GLM described above. For each subject, the physiological activity was defined by the first eigenvariate extracted from the signals from the voxels within the defined seed ROI using the VOI function implemented in SPM8. Brain regions that showed a significant effect in the interaction contrast were considered to co-vary as a functional network with the seed region more strongly in one condition over the other. This interaction contrast for every subject was used as input for the second-level random effects analysis. Participants with a higher level of integration (i.e., longer RTs due to increased lexical competition in the pause detection task for base words of trained relative to untrained novel words) might show stronger/weaker connectivity. For this reason, we also ran a second-level analysis where RTs from the pause detection task were included in the model as covariates.

Results of all second level analyses were initially thresholded at $p < .001$ (voxel-level, uncorrected). For the whole-brain search, further cluster-size statistics were used as the test statistics applying a threshold of family-wise error corrected (FWE) $p < .05$ (Hayasaka and Nichols, 2003), unless otherwise stated. We had a priori hypotheses about possible changes in the following regions: the MTL for unconsolidated and consolidated memory retrieval, left middle temporal gyrus (MTG) where phonological lexical representations are assumed to be stored (Hickok and Poeppel, 2004, 2007; Price, 2010) and mapping between word forms and word meanings takes place (Gow, 2012), and bilateral superior temporal gyrus (STG) for prelexical acoustic–phonetic representations (Gow, 2012). For these regions, we performed ROI analyses using the anatomically defined areas according to the AAL template (Tzourio-Mazoyer et al., 2002) provided by WFU PickAtlas software (Maldjian et al., 2003). For the MTL ROI, a combined region in the bilateral hippocampus and parahippocampal mask was used, and for the STG ROI, a combined bilateral STG mask was used.

Results

Behavioral data

Training

Error rates in phoneme monitoring were low ($M = 6.4\%$; $SD = 3.4\%$). No significant difference between picture-associated and form-only words was found in error rates, but responses to targets in picture-associated words were 61 ms faster (picture-associated $M =$

865 ms, $SD = 125$; form-only $M = 926$ ms, $SD = 99$, $t(23) = 6.6$, $p < .001$). This suggests that subjects successfully encoded the picture–word associations and used the picture, which appeared 200 ms before sound onset, to retrieve the word form and speed up target detection.

Recognition test (fMRI)

Word recognition in the scanner was well above chance level on both days (all conditions above 80% correct, Fig. 2A, Inline Supplementary Table S1) and there were very few false alarms on both days for both untrained novel and existing new words (all less than 3%). A repeated measures ANOVA with Day (1, 2) × Condition (picture-associated, form-only) on the percentage of hits minus false alarms revealed a main effect of Day ($F(1,23) = 27.41$, $p < .001$), indicating better performance on Day 2. A significant interaction of Day × Condition ($F(1,23) = 13.44$, $p = .001$) reflected the larger performance improvement for the picture-associated condition compared to the form-only condition (mean difference Day 2 − Day 1: picture associated Δ12.7%, form-only Δ5.8%, post-hoc paired *t*-test comparing Δ% performance for form and picture condition $t(23) = 3.67$, $p = .001$).

Inline Supplementary Table S1 can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.08.023.

For the trained novel words recognized correctly as "old", the percentage correct responses on the question whether the word was associated with a picture or not (collapsed across all confidence levels) was entered into a repeated-measures ANOVA with factors Day (1, 2) × Condition (picture-associated, form-only). This analysis showed a trend for the main effect of Day ($F(1,23) = 3.72$, $p = .066$), reflecting a small improvement on Day 2. A significant main effect of Condition ($F(1,23) = 41.73$, $p < .001$) further indicates that form-only words were responded to more accurately than picture-associated words. The interaction between Day and Condition ($F(1,23) = 19.50$, $p < .001$) was driven by the significant increase in performance on Day 2 for the picture-associated condition ($t(23) = 3.34$, $p = .006$), whereas there was

no significant change in performance for the form-only condition ($t(23) = 1.53$, $p = .282$).

Lexical competition task: pause detection

Errors and reaction times below 100 ms or above 2000 ms were excluded from analysis (3.5%). The number of errors was very low (2%). A repeated-measures ANOVA with factors Day (1, 2) × Condition (picture-associated, form-only, untrained) on the RT data (Inline Supplementary Table S1) revealed a main effect of Condition ($F(2,46) = 6.08$, $p = .005$) with form-only words being significantly slower than the untrained condition (pair-wise comparison, $p = .004$). A trend towards a main effect of Day was observed, with responses being slower on Day 2 than on Day 1 ($F(1,23) = 3.10$, $p = .091$), but no interaction between Day and Condition was observed. Because we had well-defined a priori hypotheses about the lexical competition effect on each day, we analyzed the lexical competition effect (RT difference between trained and untrained, Fig. 2B) for the picture-associated and form-only conditions separately for Day 1 and Day 2. These planned pairwise comparisons revealed a significant competition effect for form-only words on Day 2 (Day 1: $t(23) = 1.79$, $p = .086$, Day 2: $t(23) = 2.70$, $p = .013$), whereas the smaller competition effect for picture-associated words did not reach significance (Day 1: $t(23) = 0.85$, $p = .405$, Day 2: $t(23) = 1.46$, $p = .157$).

Free recall

A repeated-measures ANOVA with factors Day (1, 2) × Condition (picture-associated, form-only) showed that free recall scores were higher for picture-associated than form-only words ($F(1,23) = 8.21$, $p = .009$) and improved overnight ($F(1,23) = 48.36$, $p < .001$). On Day 1, the average recall score was 5.8% and 6.0% for form-only and picture-associated words, respectively, increasing to 29.6% for picture-associated words and 19.4% for form-only words on Day 2 (Fig. 2C, Inline Supplementary Table S1). There was a significant interaction of Day and Condition ($F(1,23) = 6.99$, $p = .015$), reflecting a larger
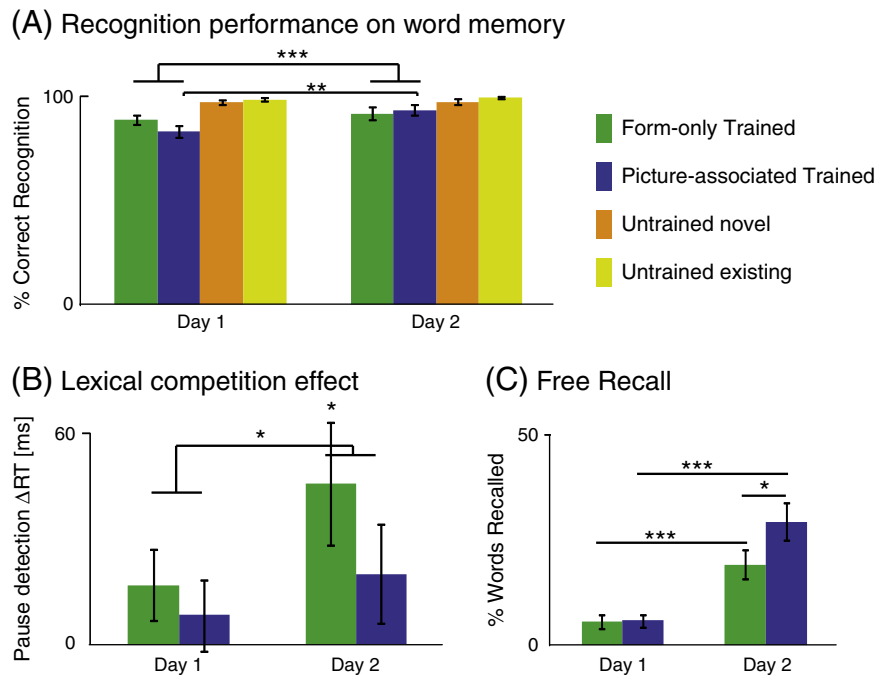
Fig. 2. Behavioral performance. (A) Recognition performance on word memory tested during the recognition test in the fMRI scanner. Percent correct responses given to each of the four memory conditions. (B) Lexical competition effect. Reaction time difference for detecting the pause between the base words of the trained novel words and untrained novel words. Reaction times were calculated from the onset of the pause until button response for pause-present stimuli, and from the time at which pause would have been inserted until button response for pause-absent stimuli. (C) Free recall. Percent number of words recalled during the free recall task. Color coding corresponds to all subpanels. *** $p < .001$, ** $p < .01$, * $p < .05$, error bar shows the standard error of means.

improvement for picture-associated words. Pairwise comparisons confirmed that there was no difference between conditions on Day 1, but on Day 2, picture-associated words were recalled more than form-only words ($t(23) = 3.03$, $p = .024$). A significant improvement from Day 1 to Day 2 in both conditions was observed (picture-associated $t(23) = 6.25$, $p < .001$; form-only $t(23) = 5.05$, $p < .001$).

Source memory test: picture–word association

Memory for the association of pictures and words was overall very good (all above 90%, Inline Supplementary Table S1). Errors for picture-associated words consisted almost exclusively of 'no picture' responses (5.2% on Day 1, 2.9% on Day 2) rather than choosing the wrong picture (0.4% on Day 1, none on Day 2). False alarms for form-only words (a choice for one of the two picture options for form-only words) also occurred infrequently (2.1% on Day 1, 3.1% on Day 2). A repeated-measures ANOVA on percent accuracy with factors Day (1, 2) × Condition (picture-associated, form-only) showed no main effect on Day or Condition, but revealed a significant interaction ($F(1,23) = 6.68$, $p = .017$). Pairwise comparisons revealed that this interaction was due to better performance on Day 1 for form-only words compared to picture-associated words ($t(23) = 2.67$, $p = .056$), whereas this difference was no longer significant on Day 2. Change within each condition from Day 1 to Day 2 was also not significant after multiple comparison corrections.

FMRI data

Activation

First we contrasted Day 1 activity for novel words that were successfully recognized as 'old', compared to untrained novel words responded to correctly as 'new'. For every participant, a contrast of (correctly recognized form-only and picture associated: hits)–(correctly recognized as new for untrained words: correct rejection) was generated and these contrast images were tested against zero using a one-sample *t*-test on the second level. This revealed multiple areas, including areas known to be involved in semantic processing (Binder and Desai, 2011; Binder et al., 2009; Price, 2010), such as the angular gyrus, MTG, and inferior- and superior frontal gyri (IFG, SFG) and an area activated for word retrieval in the middle frontal gyrus (for a review see Price, 2012) (Fig. 3A, Inline Supplementary Table S2).

Inline Supplementary Table S2 can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.08.023.

Since we had an a priori hypothesis about the involvement of the MTL on Day 1, we searched for effects within the MTL ROI. This revealed greater activity for correctly recognized trained novel words compared to correct rejection of untrained novel words in the left MTL and in the right MTL (Fig. 3D, Inline Supplementary Table S2).

On Day 2, when brain activity of successfully recognized trained words was contrasted against correctly rejected untrained words, bilateral MTL activity was still present (Figs. 3C, D, Inline Supplementary Table S2).
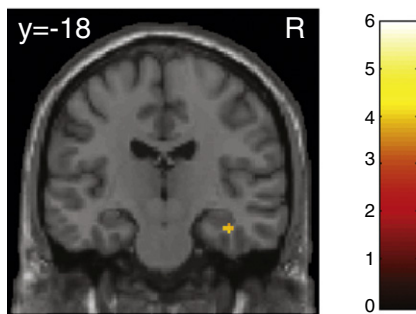

(A) Day1 Trained > Untrained novel words
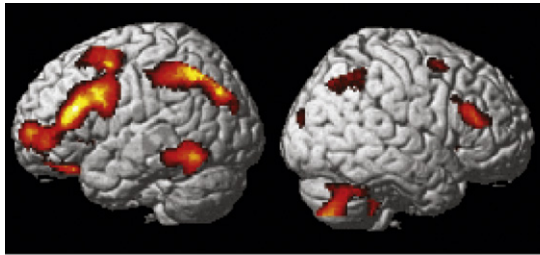
(B) Day 2 Trained > Untrained novel words
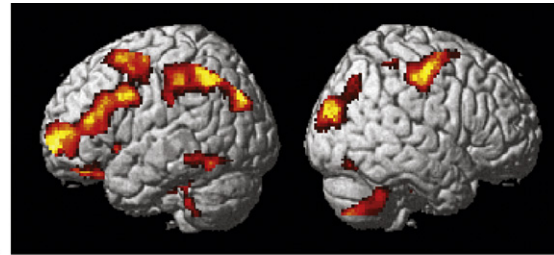
(C) Day 1 Trained > Untrained novel words

(D) Right Hippocampus (peak voxel [38 -18 -22]
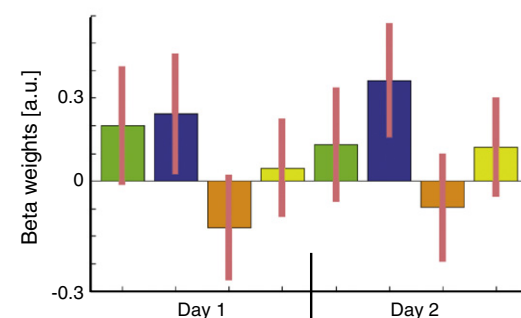
(E) Mask used for Medial Temporal Lobe ROI

Form-only Trained
Picture-associated Trained
Untrained novel
Untrained existing
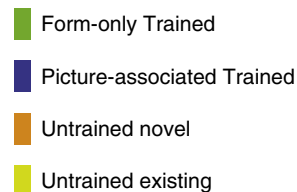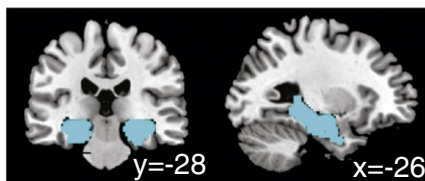
Fig. 3. Activation difference for correct recognition of novel words. (A) Significant clusters found for the trained > untrained novel words on Day 1 superimposed on a 3D rendering image provided by SPM8. (B) Significant clusters for the trained > untrained novel words on Day 2. (C) Right hippocampal involvement on Day 2 for the trained > untrained contrast is shown on the coronal slice ($P_{FWE-SVCcluster} < 0.05$ on the MTL ROI). (D) Parameter estimates for each condition extracted from the peak voxel in the right hippocampus [38 −18 −22]. (E) Mask used for the MTL ROI shown in cyan. For illustration purposes, for panels A and B, the clusters are thresholded at initial voxel level $p < .001$, with a cluster-size > 153, corresponding to $P_{FWE-cluster} < .05$. For panel C, voxels are thresholded at initial voxel level $p < .001$, with a cluster-size >44 voxels, corresponding to $P_{FWE-SVCcluster} < .05$. R = right. MTL = medial temporal lobe, a.u. = arbitrary units.

To observe activation differences between successfully recognized form-only words and picture-associated words, and how these change over a 24 hour delay, we generated contrast images for form-only hits and picture-associated hits relative to the unmodeled baseline for the two sessions. These contrast images were subjected to second-level comparisons using ANOVAs with factors Day (1×2) and Condition (form-only × picture-associated). This revealed a main effect of Day with Day 1 > 2 in the bilateral striatum (pallidum/putamen/caudate, Fig. 4A) and Day 2 > 1 in multiple regions including the bilateral STG, bilateral rolandic operculum, bilateral posterior MTG, left IFG (pars triangularis), left middle frontal gyrus, and left (and a subthreshold cluster in the right) insula (Fig. 4B, Table 1). An interaction effect showed a cluster in the left posterior MTG (Fig. 4C) when the left MTG mask was applied (Fig. 4F). Post-hoc analyses showed that this interaction in the left MTG was driven by form-only words showing higher activity than picture-associated words on Day 1, and the picture-associated words showing an increase in activity on Day 2 compared to Day 1 (Fig. 4E).

ANOVA analysis might not be sensitive enough for subtle effects when two different sessions are compared against each other, even if we control for baseline differences. For completeness we also performed comparisons between conditions and sessions separately, by generating a condition- or session-specific contrast on the single subject level. We then subjected these contrast images to a second-level group analysis, performing a one sample t-test on each of the contrasts. First we compared Day 1 activity elicited by hits for form-only and picture-associated novel words, in order to assess potential differences between conditions that might have existed before consolidation. Several areas showed increased activation for form-only relative to picture-associated words, including the bilateral MTG, bilateral postcentral gyrus extending to supramarginal gyrus, left middle section of STG/MTG, and right postcentral gyrus extending to superior frontal gyrus/superior medial frontal gyrus/IFG (Fig. 5A, Table 2). No areas survived the cluster-level threshold for the reverse contrast of enhanced activation of picture-associated relative to form-only words.
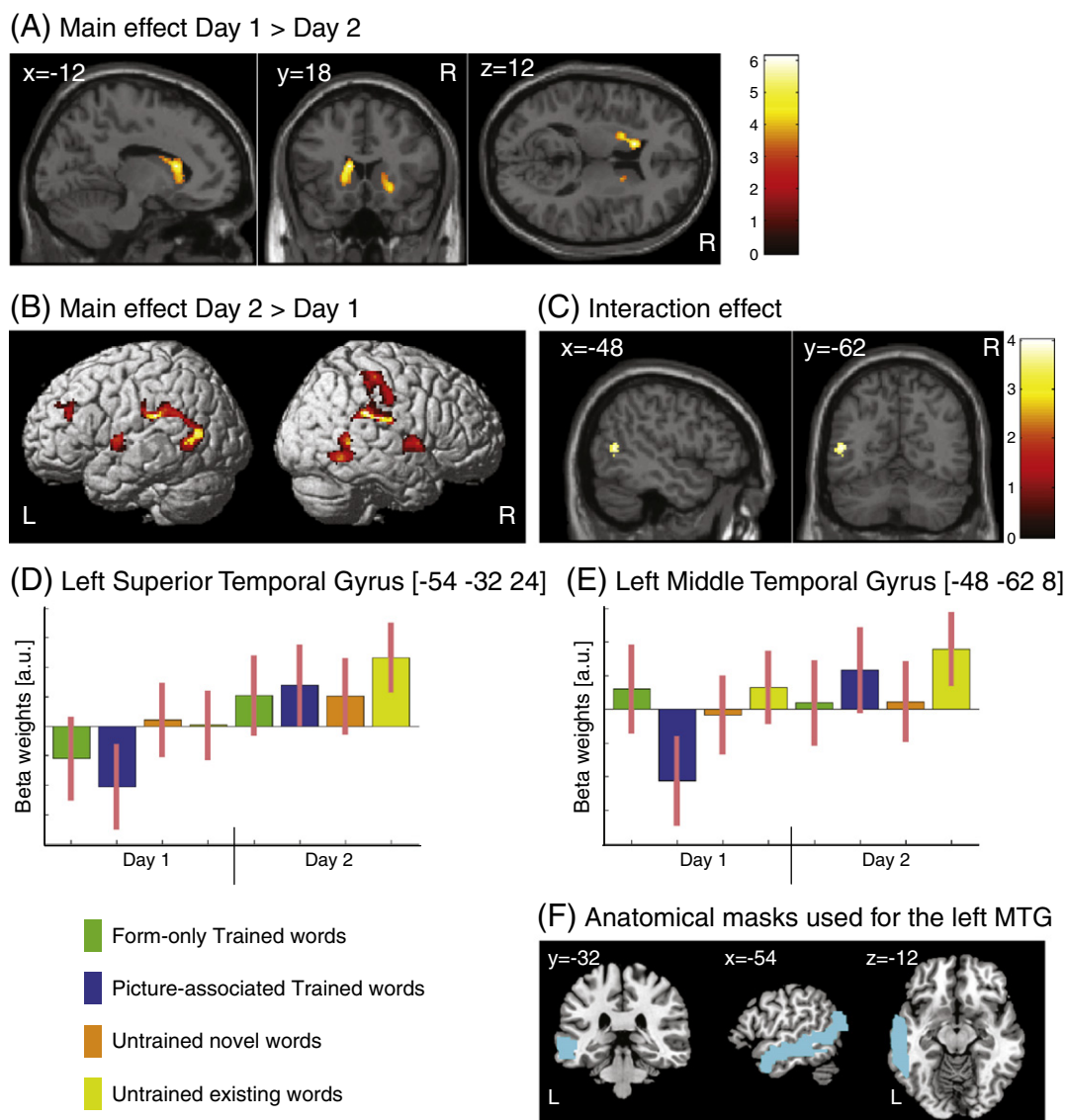
Fig. 4. Activation change over time for correctly recognized words. (A) Main effect of Day 1 > Day 2 activation superimposed on a canonical T1 image provided by SPM8. Bilateral striatal activation is larger on Day 1 than Day 2. (B) Main effect of Day 2 > Day 1 superimposed on a 3D rendering image provided by SPM8. (C) An interaction effect was found when the left MTG ROI was applied. For illustration purposes, for panels A and B the clusters are thresholded at initial voxel level $p < .001$, with a cluster-size >155, corresponding to $P_{FWE-cluster} < .05$ and C with a cluster size >43 corresponding to $P_{FWE-SVCcluster} < .05$. (D) Parameter estimates of the peak voxel in the left superior temporal gyrus [$-54 -32 24$] for each of the conditions. (E) Parameter estimates of the peak voxel in the left middle temporal gyrus (MTG) [$-48 -62 8$] for each of the conditions. (F) The anatomical left middle temporal gyrus mask used for the ROI analysis. L: left, R: right, a.u. = arbitrary units.

Table 1

Main effect Day 1 > Day 2

Cluster		Area	MNI coordinates at local maxima			
P_{FWE}	Size		x	y	z	t value
0.001	562	L caudate	−12	18	12	6.13
		L putamen	−20	12	−2	5.21
0.004	404	R putamen	22	14	−4	4.51
		R caudate	18	22	4	3.49

Main effect Day 2 > Day 1

Cluster		Area	MNI coordinates at local maxima			
P_{FWE}	Size		x	y	z	t value
0	803	R rolandic operculum	48	−26	24	6.17
		R postcentral gyrus	50	−28	56	5.29
		R postcentral gyrus	56	−22	46	4.61
0.125	158	L rolandic operculum	−50	−2	4	5.31
		L insula	−38	−14	−4	3.64
0.073	192	R anterior superior temporal gyrus	50	−2	0	5.26
		R insula	38	−14	−4	3.37
0.012	315	R middle temporal gyrus	50	−50	6	4.99
		R inferior temporal gyrus	58	−56	−10	4.3
		R posterior middle temporal gyrus	66	−50	−6	4.18
0	698	L posterior superior temporal gyrus	−54	−32	24	4.92
		L supramarginal gyrus	−60	−44	26	4.85
		L posterior middle temporal gyrus	−58	−60	0	4.85
0.026	259	L inferior frontal gyrus (pars triangularis)	−34	34	26	4.04
		L middle frontal gyrus	−26	42	32	3.81

Interaction effect

Cluster		Area	MNI coordinates at local maxima			
P_{FWE}	Size		x	y	z	t value
0.04[a]	64	L posterior middle temporal gyrus	−48	−62	8	4.02
		L posterior middle temporal gyrus	−44	−56	10	4
		L posterior middle temporal gyrus	−54	−60	2	3.32

Cluster size is the number of voxels (size $2 \times 2 \times 2$ mm) that comprises the cluster. Activations are thresholded at cluster-size significant at family-wise error corrected $p < .05$ (>139 voxels), initial voxel level threshold at $p < .001$. L: left, R: right.

[a] Small volume correction on anatomically defined medial temporal lobe mask (extent threshold cluster-size > 54 corresponding to $P_{FWE} < .05$).

On Day 2, no clusters showing greater activation for form-only words compared to picture-associated words that survived the threshold were observed, although one cluster in the bilateral anterior cingulate cortex (ACC; $p = .088$, $k = 140$, peak voxel [−6 24 20]) showed a trend. The reverse contrast of picture-associated > form-only words on Day 2 revealed no significant clusters above threshold.

If the associated picture memory was intact for picture-associated words, however, extra activation pertaining to visual memory might be present. To test this, we modeled the confidence level of the picture memory response as a parametric modulator on the single subject GLM analysis and compared the successful recognition of picture-associated words to form-only words controlling for picture memory strength. This contrast revealed the left anterior cingulate cortex extending to left middle orbital gyrus and right superior orbital gyrus, and a left MTG cluster extending to fusiform gyrus, inferior temporal gyrus and inferior occipital gyrus (Fig. 5B, Inline Supplementary Table S3). This suggests that when strength of the picture memory was controlled, the recognition of picture-associated words revealed increased activation in areas known to be involved in visual processing related to semantic information, such as the left MTG-fusiform areas (Binder and Desai, 2011; Price, 2010). Within the anatomical mask of the MTL ROI, one cluster showed a direction towards higher activation for picture-associated larger than form-only condition, although it did not survive our statistical threshold (Inline Supplementary Table S3). This hints at a sustained role of the MTL at the time of retrieval for picture-associated words compared to form-only words on Day 2 if the memory of the paired picture was present (Fig. 5B right).

Inline Supplementary Table S3 can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.08.023.

Memory strength might have decayed due to forgetting over the course of 24 h. For this reason, we also modeled the confidence rating on the word memory as parametric modulation on the single subject GLM analysis. The pattern observed, of activity being greater for picture-associated than for form-only words, was very similar to the results obtained when we controlled for the memory confidence strength for the pictures (see above). The reverse pattern (form-only > picture-associated) did not show any significant clusters above threshold.

We next compared changes in activity from Day 1 to Day 2 separately for form-only words and picture-associated words using one-sample t-tests at the group level. Similar to the main effect of Day found in the ANOVA contrast reported above (Fig. 4A), both the form-only and the picture-associated conditions revealed more activity in the striatum for Day 1 than Day 2, although the picture-associated condition was just below our threshold. An activation increase from Day 1 to Day 2 for picture-associated words was found in the left inferior parietal lobe extending to the angular gyrus, supramarginal gyrus, rolandic operculum, STG, MTG, Heschl's gyrus, and insula, right STG extending to the insula and precentral gyrus, right inferior temporal gyrus/MTG, right supplementary motor area extending to the middle cingulate cortex, and bilateral middle and superior frontal gyri (Fig. 5C, Inline Supplementary Table S4). The comparison of Day 2 versus Day 1 activation for form-only words did not reveal any clusters above threshold, although bilateral clusters in posterior STG, close to Heschl's gyrus were observed at the subthreshold level (local maxima [−36 −34 18], left cluster size

(A) Day 1 Form-only > Picture-associated novel words

(B) Day 2 Picture-associated > Form-only novel words modulated by the strength of picture memory

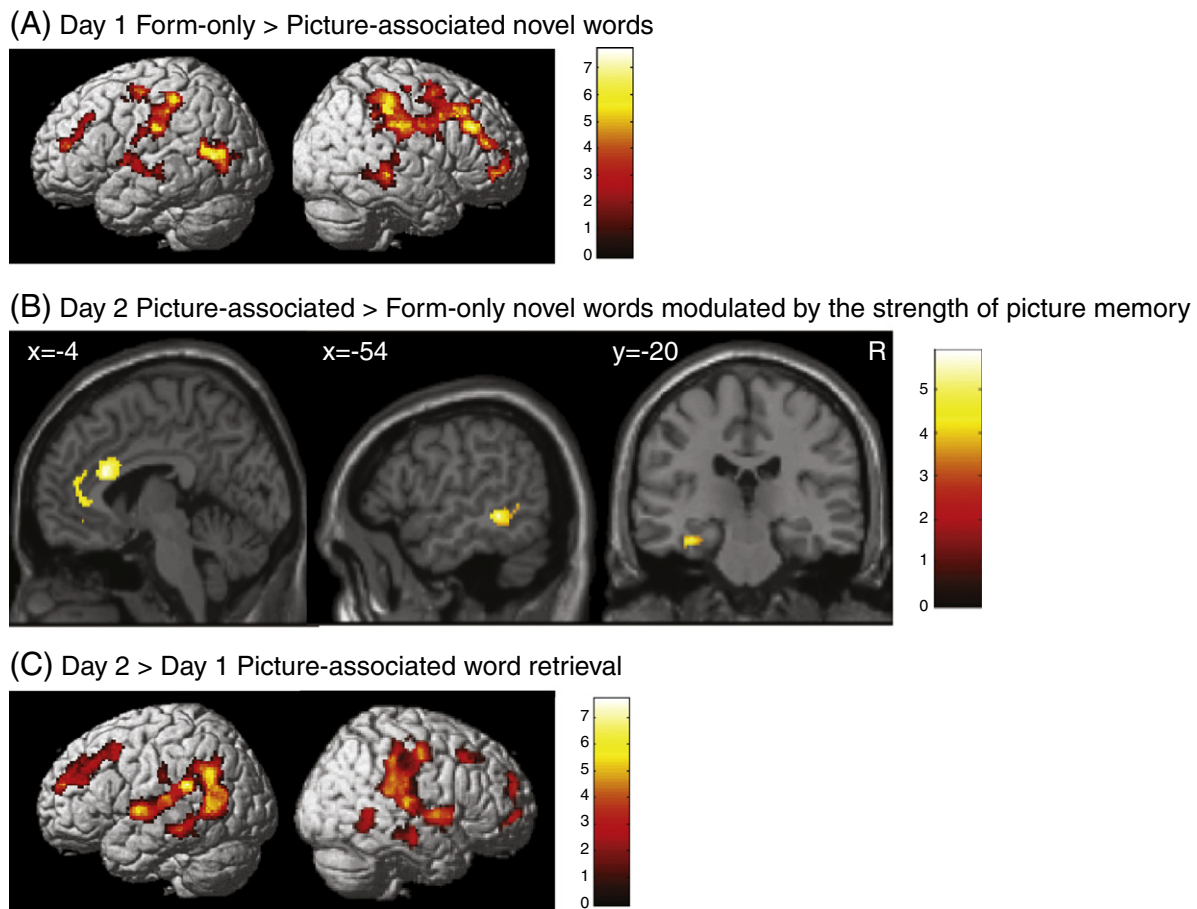### (C) Day 2 > Day 1 Picture-associated word retrieval

Fig. 5. Activity differences between conditions in each of the sessions. (A) Day 1 comparison, form-only > picture associated words. (B) Day 2 comparison, picture-associated > form-only words when the strength of the picture memory was controlled for. (C) Day 2 > Day 1 comparison for picture-associated words. For illustration purposes, the clusters are thresholded at initial voxel level $p < .001$, with a cluster-size > 155, corresponding to $P_{FWE-cluster} < .05$, except for panel B right where the cluster is thresholded at initial voxel level $p < .001$, with a cluster-size >23, corresponding to $P_{FWE-cluster} < .106$.

$k = 81$, $p = .281$, right local maxima $[48 - 26\ 16]$, $k = 50$, $p = .599$). These clusters overlapped with those that exhibited an overnight increase in activation for picture-associated words. This suggests that the main effect of Day $(2 > 1)$ in the above ANOVA analysis (Fig. 4B) was mainly driven by the changes for the picture-associated words.

Inline Supplementary Table S4.

Inline Supplementary Table S4 can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.08.023.

Connectivity

The old/new recognition task that participants performed in the scanner required matching aurally presented stimuli to existing memory traces. We assumed therefore that activity in the auditory cortex in the STG would have strong functional connectivity with areas supporting word storage and retrieval. The areas involved in these processes were expected to exhibit an overnight shift, potentially modulated by the training condition (form-only versus picture-associated). To test this, we performed a PPI analysis for each day, with the seed region defined as the portion of the bilateral STG in which trained novel words elicited activation compared to baseline (for details, see Methods section). We searched for functional connectivity that was greater for the form-only than for the picture-associated condition and vice versa, separately for each day.

On Day 1, PPI analysis revealed no regions above threshold on the whole brain level for both directions (form-only > picture-associated, picture-associated > form-only). Since we had an a priori hypothesis about the involvement of the MTL in binding multimodal associations, we performed a small volume correction (SVC) on the anatomical

MTL ROI. This analysis revealed that successfully recognized picture-associated words connected more to the right hippocampus than the form-only words (local maxima $[28 - 12 - 28]$, $k = 101$, $P_{SVC-FWE} = .006$).

Because the STG is known to play a role in the perception and recognition of speech sounds (Hickok and Poeppel, 2007), we also searched within a bilateral anatomical STG ROI. Form-only words showed a stronger connection with the right STG (local maxima $[54 - 6 - 8]$, $k = 50$, $P_{SVC-FWE} = .033$). Fig. 6A illustrates these contrasts (red = seed ROI, blue = stronger for picture-associated condition, green = stronger for form-only condition).

On Day 2, no clusters survived the whole brain comparison for either the form-only greater than picture-associated condition or the reverse contrast. However, we may expect that if the behavioral lexical competition effect in the pause-detection task (i.e., the difference in RT between base words of the trained (novel-competitor) and untrained (control) novel words) reflected the degree of integration, the magnitude of this effect might correlate with the degree of functional connectivity between auditory processing areas and phonological lexical representational areas such as the left posterior MTG (Gow, 2012; Hickok and Poeppel, 2004, 2007) and the posterior STG where prelexical phonological processing is supposed to take place (Gow, 2012; Graves et al., 2008). We tested for areas in which connectivity differed between the form-only and picture-association conditions, and in which this difference correlated with the behavioral competition effect. We added two covariates to the model, containing the magnitude of the lexical competition effect on Day 2 (i.e., the RT difference between base words of trained and untrained novel words), one for the form-only condition and another

Table 2
Activation contrast for form-only > picture-associated novel words remembered on Day 1.

Cluster		Area	MNI coordinates at local maxima			
P_{FWE}	Size		x	y	z	t-Value
0	1715	L precentral	−30	−6	54	7.71
		L supramarginal gyrus	−52	−24	42	6.3
		L postcentral	−56	−20	32	6.04
0	894	L middle temporal gyrus	−46	−60	10	7.21
		L middle occipital gyrus	−38	−70	16	5.47
0	3308	R inferior frontal lobe (pars triangularis)	44	28	30	6.99
		R middle frontal gyrus	36	32	32	6.25
		R postcentral	48	−30	54	6.21
		R superior frontal gyrus	16	−6	62	6.03
		R superior medial gyrus	62	−20	30	5.94
0.004	245	L middle frontal gyrus	−38	46	16	5.87
0	349	R posterior superior temporal gyrus	62	−32	2	5.51
		R middle temporal gyrus	52	−32	−8	4.82
		R middle temporal gyrus	62	−32	−12	4.2
0.004	248	L middle superior temporal gyrus	−52	−8	0	4.75
		L middle temporal gyrus	−58	−22	−2	4.58
		L rolandic operculum	−46	0	8	4.17

Cluster size is the number of voxels (size 2 × 2 × 2 mm) that comprises the cluster. Activations are thresholded at cluster-size significant at family-wise error corrected $p < .05$ (>139 voxels), initial voxel level threshold at $p < .001$. L: left, R: right.

for the picture-associated condition. The strength of the functional connectivity (form-only > picture associated) between auditory cortex and a posterior MTG cluster showed a positive correlation with the competition effect for the form-only condition (Fig. 6B, peak local maximum [−54 −62 4], within the left MTG ROI mask $P_{SVC-FWE} = .048$). This suggests that those participants who showed a greater competition effect in the form-only condition had stronger functional connectivity between the auditory seed region and the left posterior MTG when successfully recognizing form-only words compared to picture-associated words. This cluster overlapped with the area that showed greater activation for recognizing picture-associated words on Day 2 compared to Day 1 (Fig. 5C). This effect was not found in the STG ROI. No correlations with the competition effect were found for the reverse contrast (the picture-associated condition > the form-only condition).

Discussion

In the current study, we investigated how novel words are encoded and integrated into the existing lexicon and how this is expressed in the neural networks involved in storing and retrieving memory traces immediately after learning and following a period of overnight consolidation. In particular, we tested the complementary learning systems account of lexicalization (Davis and Gaskell, 2009; McClelland et al., 1995) by asking if richness of information at encoding has an effect on both lexical integration and memory retrieval.

Lexical integration

The difference in pause detection between existing words that were phonological neighbors of trained novel words and those that were neighbors of untrained novel words (control words) suggests that trained novel words were competing with existing words during spoken-word recognition. The presence of this competition effect on Day 2, but not on Day 1, once more provides empirical support for the role of (sleep) consolidation in the lexicalization of novel words (Davis and Gaskell, 2009).

A novel finding of this study was that, after 24 h, the lexical competition effect had emerged in the form-only condition but was not yet
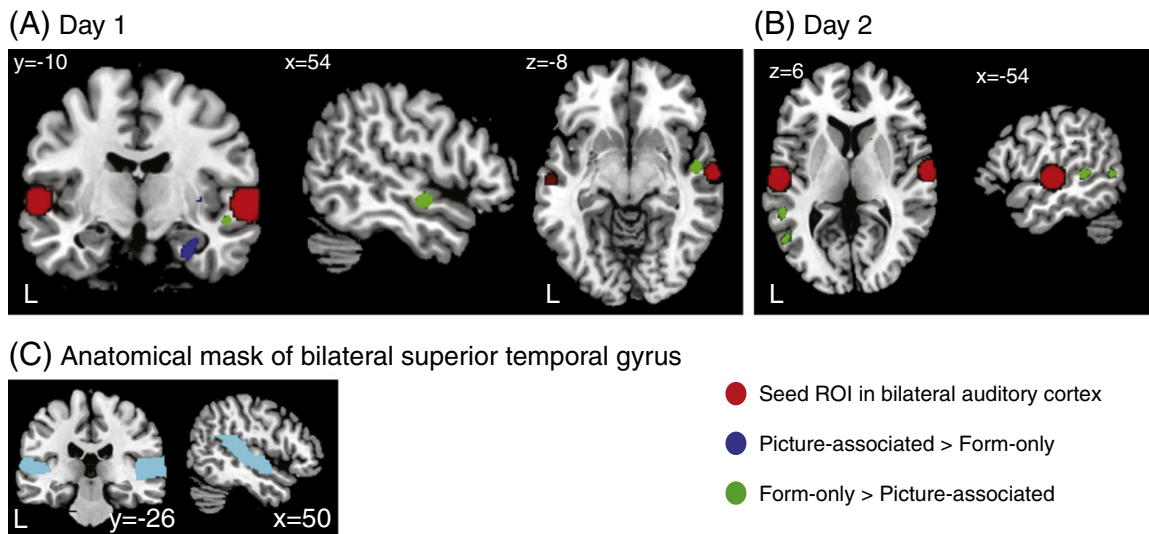

(A) Day 1
y=-10   x=54   z=-8

(B) Day 2
z=6   x=-54

(C) Anatomical mask of bilateral superior temporal gyrus
y=-26   x=50

● Seed ROI in bilateral auditory cortex
● Picture-associated > Form-only
● Form-only > Picture-associated

Fig. 6. Functional connectivity from the bilateral auditory cortex. (A) Day 1. (B) Functional connectivity strength (form-only > picture-associated words) that correlated with the lexical competition effect on Day 2 in the left posterior middle temporal gyrus. Red: seed region in auditory cortex, blue: picture-associated > form-only, green: form-only > picture-associated. Results of functional connectivity analyses are shown as significant after small volume correction on the anatomical regions of interest. (C) The anatomical mask of the bilateral superior temporal gyrus used for small volume correction on the Day 1 contrast. Please refer to Fig. 3E for the medial temporal lobe mask and Fig. 4F for the left middle temporal lobe mask. L: left.

apparent for the picture-associated words. We suggest that this result is due to the differences in the richness of information that was available during training between the form-only and the picture-associated novel words. It was possible for the learners to form stronger associations between the novel form-only words and their existing phonological neighbors than between the picture-associated words and their neighbors because there was no visual information for the form-only words. That is, learners had only the phonology of the form-only words to build their memory on, while, in contrast, they could link the picture-associated words also to visual cues (i.e., they could build rudimentary meanings for these words). In line with the complementary learning systems account, stronger associations between the phonological forms of novel and existing words should facilitate the interleaving of existing and novel information into neocortical memory, and hence the stronger phonological competition effect found for the form-only words.

An alternative possible outcome was exactly the opposite of that found: more lexicalization for the picture-associated words than for the form-only words. Stronger evidence of lexicalization (as measured in a perceptual learning task) has previously been found when novel words were associated with meanings (in definitions and in pictures) than when they were not (Leach and Samuel, 2007). What might account for this difference between studies? One possibility is simply that the two studies used different measures of lexical integration. There are different components of lexicalization (e.g., integration of the word's form into the lexicon, of its morphosyntactic properties and of its meaning). These components may be measured to different degrees by different tasks, and the richness of information in the input may contribute differentially to these components. It is therefore important to emphasize that we are not arguing that information about novel word meanings (e.g. from pictures) necessarily inhibits word learning. It is obvious that new words, in natural learning situations, must have meanings, and it remains plausible, in spite of the current results, that adding meanings to words is likely to enhance components of lexicalization. What we have shown, however, is that additional pictorial information can delay one of these components: integration of the form of a new word into the existing phonological lexicon that is used during spoken-word recognition.

Our findings on Day 1 are not consistent with previous work using the pause detection task (Davis et al., 2009; Dumay and Gaskell, 2007; Gaskell and Dumay, 2003; Tamminen and Gaskell, 2008; Tamminen et al., 2010) in that form-only words already showed a trend towards interference on Day 1. At this point we can only speculate that this inconsistency is due either to a difference in the tasks that were used (in former studies participants were exposed only to word forms, whereas half of our novel words were associated with pictures) or to the bilingual nature of the participant population (our Dutch participants have learned multiple languages throughout their education) as opposed to the monolingual nature of most of their peers in English-speaking countries (where the previous studies have been conducted). Furthermore, the training session was long (on average 75 min) and participants were exposed to the novel words multiple times, creating a possible interleaved learning opportunity during the training phase of the experiment. Consistent with this possibility, Lindsay and Gaskell (2013) have recently reported that lexical competition (as measured using the pause detection task) can emerge in the absence of sleep in the context of an interleaved training protocol.

In line with the proposal that form-only words are at an advantage in lexical form integration, the fMRI results on Day 1 already showed an activation difference between the form-only and picture-associated conditions when words were successfully recognized. Form-only words elicited more neocortical activation than picture-associated words, especially in the areas known to reflect lexical access processes in the posterior STG (Gow, 2012; Graves et al., 2007, 2008), and those interfacing phonology with the semantic lexicon in the left posterior MTG (Gow, 2012; Hickok and Poeppel, 2004, 2007). When pseudowords are very similar to an existing word, processing of pseudowords can activate brain areas related to real word processing (Raettig and Kotz, 2008) probably due to incidental co-activation of the similar existing words. Our finding of increased neocortical activity during recognition of form-only words on Day 1 may reflect the co-activation of phonological neighbor words which were already present in the neocortical lexical/semantic memory system.

Functional connectivity from the auditory cortex also revealed a stronger connection to the STG for form-only words relative to picture-associated words on Day 1. Although the function of the right STG in language processing is reported less often, it has been found to be activated during auditory sentence comprehension (Crinion and Price, 2005), and selectively for words compared to tones (Binder et al., 2000). This suggests that the right temporal lobe may also have a role in semantic processing. In a similar study to ours, Davis et al. (2009) found stronger activation for unconsolidated novel words during the pause detection task compared to consolidated novel words or existing words. They interpreted less activation for the consolidated words as a sign of sharper, better-tuned neural representations. On the other hand, the STG is also known to increase its activity with a higher demand on the lexical access system (Newman and Twieg, 2001). From this view one could speculate that in our case, recognition of form-only words required extra effort in lexical access to their existing phonological neighbor words on Day 1.

The activation pattern is consistent with the idea that when subjects recognized the form-only words, phonological representations of their existing neighbor words were also activated causing an increase in activation in areas that support semantic/lexical processing. This co-activation suggests that a strong link between novel and existing word representations started to emerge during or immediately after encoding. Such partial integration on Day 1 may facilitate the overnight consolidation process, leading to the emergence of significant lexical competition on Day 2. The positive correlation between the lexical competition effect on Day 2 and functional (auditory cortex–posterior MTG) connectivity strength for form-only words relative to picture-associate words endorses the idea that a tight link between existing words and newly learned words is a key factor contributing to the emergence of a lexical competition effect in the form-only condition. In other words, the strong link between the novel word and its phonological neighbor base word helped the retention of the word forms, leading to correct recognition on Day 2, and caused stronger interference during the processing of the similar existing words.

The memory retrieval network

Memory strength was stronger for the picture-associated words on Day 2, as measured by the number of freely recalled words. In order to be able to retrieve from memory without any cues, the memory trace needs to be strong, whereas recognizing an item as old can be performed even with a vague familiarity notion. Thus free recall is thought to be a more sensitive measure for the strength of memory traces than recognition accuracy. It might seem strange that memory performance increased across days when normally declarative memories can only decay with time. The memory performance increase from Day 1 to Day 2 might be affected by the fact that participants were exposed to the words again after the free recall session on Day 1, during the source memory test and during the recognition test on Day 2. Importantly, however, the difference we observed between the number of words recalled in the form-only and picture-associated conditions on Day 2 cannot be due to the extra exposure in these tests on Day 1, since the amount of extra exposure was the same for the two conditions. We suggest that the superior memory performance for the picture-associated words arose because words in this condition could benefit more from the joint availability of episodic and semantic memory systems than the form-only words.

Evidence for this claim comes from both the activation data and the connectivity analyses. We observed an increase in activity in the left posterior MTG for picture-associated words from Day 1 to Day 2. From

the complementary learning systems account, we could speculate that this increase in activity could be interpreted as the emergence of neocortical representations for the phonological form of picture-associated words. Rather than storing semantic information per se, the posterior MTG has been suggested to store lexical representations which bind phonological information stored in the STG, via amodal representations in the temporal pole, to widely distributed, conceptual representations (Gow, 2012; Hickok and Poeppel, 2004, 2007). This distinction between semantic and lexical storage is supported by the dissociation between damage to the posterior MTG, which is associated with relatively intact semantic representations and impaired word retrieval, whereas damage to the temporal pole appears to impoverish semantic representations themselves (Patterson et al., 2007). The present result fits well with such a two-step model of lexical access, which predicts that consolidated novel words should activate lexical representations in the left posterior MTG. Furthermore, we observed an increase in activation in the STG as well. With the emergence of the novel word's phonological lexical entry on Day 2, weak (i.e., not statistically significant) lexical competition with the existing neighbor words also arose for picture-associated words on Day 2, resulting in a more effortful process reflected in the STG. Our functional connectivity data on Day 2 for form-only words additionally show that access to a lexical representation (either of the novel word or the phonological neighbor word, when cued with the novel word) in the posterior MTG is reflected in the extent to which the novel word is able to inhibit phonological neighbors in the lexical competition task.

As for the episodic memory system, in line with the complementary learning systems account, the MTL showed stronger activation for recognition of the trained words in both training conditions (collapsed across the picture-associated and form-only conditions) on Day 1 compared to untrained condition. Unlike in the Davis et al. (2009) study, though, MTL activity was still visible for words that were successfully recognized on Day 2. When recognition of picture-associated words with intact picture memory was contrasted against form-only words on Day 2, a subthreshold cluster in the left MTL region was observed (Inline Supplementary Table S3), suggesting a role of MTL in successful retrieval of auditory–visual associations.

To what extent, therefore, is the episodic memory system involved in the retrieval of newly learned words? The hippocampus is thought to be involved in binding memory traces that are distributed across multiple brain areas (Davachi, 2006; Eichenbaum and Cohen, 2001; Konkel and Cohen, 2009; Mayes et al., 2007). In our case, binding of visual information and phonological word-form representations was required for picture-associated words during the training phase. Our hypothesis that the hippocampus initially binds the auditory and visual information was partially confirmed with the connectivity analysis. Picture-associated words showed stronger connections between the auditory cortex and the MTL than the form-only words did. Although this functional connectivity difference between the two conditions was no longer observable on Day 2, vivid picture memory still seemed to activate the MTL region (despite the small effect), suggesting a sustained involvement of the hippocampal system even after a night of sleep for the picture-associated words with intact picture memories. It is possible that richer vivid memories reside in the hippocampal network even after a passage of time (Winocur and Moscovitch, 2011). We thus suggest that picture-associated words benefit from the simultaneous support of two memory systems, the hippocampal episodic system (MTL) and the neocortical lexical system (posterior MTG), which aids the retrieval process and leads to superior memory performance on Day 2 even after the memory trace may have started to decay in the episodic memory system.

We also suggest, however, that this complementary learning systems account is consistent with the absence of statistically reliable lexical competition for the picture-associated words on Day 2. One possibility is that, if two new pieces of information (i.e., a new phonological word-form and its unfamiliar visual referent picture), both not yet existing in the neocortical semantic memory network, need to be incorporated into that system, more repetition or time may be required than when only one piece of information is added to the existing lexicon (i.e., the associative link between the novel word form and an already existing phonological neighbor word). Another possibility, as we suggested earlier, is that associations between picture-associated words and their existing phonological neighbor words may be weaker than those for the form-only words. Thus, while picture-associated words are at an advantage in explicit memory tests because of the existence of the two memory systems, form-based words are at an advantage in the integration into the phonological lexicon, with a strong link to the phonological neighbor words, because of the way information is transferred from the episodic system to the lexical system.

Comparison with previous fMRI findings on lexicalization

Unlike the reduced STG activity for consolidated words reported by Davis et al. (2009), we observed an increase in STG activation on Day 2 overall, but more strongly in the picture-associated condition. Changes in the experimental setup could be a reason for this difference. As the STG has been shown to be activated when a higher demand is placed on the phonological lexical access system, such as in a phoneme monitoring task (Newman and Twieg, 2001) or during production of low frequency words (Graves et al., 2007), the activity change may reflect the acoustic and phonological analysis process or an effortful search process to reach a lexical entry. Davis et al. (2009) argue that similarity in neural responses to consolidated novel words and existing words implies cortical integration for the consolidated novel words, and therefore fewer resources are needed for phonological processing. However, their use of a pause-detection task during scanning to measure responses to novel words may have caused subjects to devote more resources to acoustic and phonological analysis than lexical retrieval. Although the difference they report may be caused by changes in the words' representations through consolidation, their result may still reflect the executive processing difference as a consequence of this change. The old/new recognition task employed in the present study, in contrast, required participants to match the auditory input to an internal memory trace. Activation increases would be expected in areas such as the left posterior MTG if that is where the phonological lexicon is represented (Gow, 2012; Hickok and Poeppel, 2004, 2007), at least once those representations have been established for novel words. Furthermore, stronger competition-related activity would be expected in areas such as the STG because of lexical access attempts for the entries of both the novel word and its existing phonological neighbor word.

In line with the complementary learning systems account that newly learned information will be integrated into the neocortical structures with consolidation, we observed increased neocortical activity on Day 2. Interestingly, the increase in the neocortical involvement after 24 h was greater and more widespread for picture-associated than form-only words. It is plausible that having more than just phonological information about a word would involve neocortical areas beyond those supporting phonological representation.

Instead of the two training sessions in Davis et al. (2009), we opted for one training session and two test sessions. In this way, we could track changes in the retrieval network over time for the same stimuli. This design however did not allow us to dissociate whether the retrieval process on Day 2 is purely the result of consolidation or whether it might have been influenced by the extra exposure to the stimuli during the tests on Day 1. Despite this concern, since participants were exposed to the novel words much more extensively during the initial training than additional exposure during the test, the change is more likely to be due to consolidation than to more exposure. Furthermore, the difference between the two conditions on Day 1, and how that differed on Day 2, reflects to some extent the consequence of overnight consolidation, as the number of exposures, both initially and between the two test sessions, was equal for the two conditions.

Conclusion

In this study, we aimed to elucidate the neural correlates of phonologically learned novel words and how they change with an overnight delay. Additionally, we investigated whether there is a different consolidation trajectory in the presence/absence of extra visual information associated with the new words. Immediately after training, form-only words elicited stronger frontal and temporal activation than picture-associated words, suggesting that a link with the cortically represented lexical network was already recruited on Day 1. Form-only words furthermore showed stronger coupling between the auditory cortex and the superior temporal gyrus, possibly reflecting deeper phonological processing or stronger associations with the existing phonological lexicon. We have argued that these stronger links between new and old words made it easier for listeners to integrate the new words into the phonological lexicon, as demonstrated by the lexical competition effect on Day 2. Furthermore, the strength of the functional connectivity on Day 2 between the auditory and left posterior middle temporal cortices was predicted by the degree to which form-only words competed during the processing of their base words. Picture-associated words, in contrast, showed better memory retention on Day 2. On the neural level, a stronger involvement of the hippocampal memory system both on Day 1 and Day 2 was observed for this condition, with an increased neocortical activation on Day 2, suggesting that two memory systems aided the retrieval of picture-associated words on Day 2. Together, these results imply that the ability to retrieve novel words from memory is dependent on the memory systems available at retrieval, that a tight connection between novel and existing representations of phonological form underlies the emergence of lexical competition effects, and that the success of the integration process depends on the richness of the information that is available during the encoding of the novel word forms.

Acknowledgments

Conflict of interest

The authors have declared that no competing interests exist.

References

Alvarez, P., Squire, L.R., 1994. Memory consolidation and the medial temporal lobe: a simple network model. Proc. Natl. Acad. Sci. U. S. A. 91 (15), 7041–7045.

Baayen, R.H., Piepenbrock, R., van Rijn, H., 1993. The CELEX Lexical Database (CD-ROM). L. D. Consortium. University of Pennsylvania, Philadelphia, PA.

Binder, J.R., Desai, R.H., 2011. The neurobiology of semantic memory. Trends Cogn. Sci. 15 (11), 527–536.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10 (5), 512–528.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex 19 (12), 2767–2796.

Boersma, P., Weenink, D., 2012. Praat: Doing Phonetics by Computer [Computer program]. Version 5.3.23. from www.praat.org/.

Carey, S., Bartlett, E., 1978. Acquiring a single new word. Proc. Stanf. Child Lang. Conf. 15, 17–29.

Crinion, J., Price, C.J., 2005. Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. Brain 128 (12), 2858–2871.

Davachi, L., 2006. Item, context and relational episodic encoding in humans. Curr. Opin. Neurobiol. 16 (6), 693–700.

Davis, M.H., Gaskell, M.G., 2009. A complementary systems account of word learning: neural and behavioural evidence. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364 (1536), 3773–3800.

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

Dumay, N., Gaskell, M.G., 2007. Sleep-associated changes in the mental representation of spoken words. Psychol. Sci. 18 (1), 35–39.

Dumay, N., Gaskell, M.G., 2012. Overnight lexical consolidation revealed by speech segmentation. Cognition 123 (1), 119–132.

Eichenbaum, H., Cohen, N.J., 2001. Conditioning to Conscious Recollection: Memory Systems of the Brain. Oxford University Press, New York.

Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6 (3), 218–229.

Gaskell, M.G., Dumay, N., 2003. Lexical competition and the acquisition of novel words. Cognition 89 (2), 105–132.

Gitelman, D.R., Penny, W.D., Ashburner, J., Friston, K.J., 2003. Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. NeuroImage 19 (1), 200–207.

Gow, D.W., 2012. The cortical organization of lexical knowledge: a dual lexicon model of spoken language processing. Brain Lang. 121 (3), 273–288.

Graves, W.W., Grabowski, T.J., Mehta, S., Gordon, J.K., 2007. A neural signature of phonological access: distinguishing the effects of word frequency from familiarity and length in overt picture naming. J. Cogn. Neurosci. 19 (4), 617–631.

Graves, W.W., Grabowski, T.J., Mehta, S., Gupta, P., 2008. The left posterior superior temporal gyrus participates specifically in accessing lexical phonology. J. Cogn. Neurosci. 20 (9), 1698–1710.

Hayasaka, S., Nichols, T.E., 2003. Validating cluster size inference: random field and permutation methods. NeuroImage 20 (4), 2343–2356.

Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92 (1–2), 67–99.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8 (5), 393–402.

Konkel, A., Cohen, N.J., 2009. Relational memory and the hippocampus: representations and methods. Front. Neurosci. 3.

Leach, L., Samuel, A.G., 2007. Lexical configuration and lexical engagement: when adults learn new words. Cogn. Psychol. 55 (4), 306–353.

Lindsay, S., Gaskell, M.G., 2013. Lexical integration of novel words without sleep. J. Exp. Psychol. Learn. Mem. Cogn. 39 (2), 608–622.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. NeuroImage 19 (3), 1233–1239.

Martin, A., Chao, L.L., 2001. Semantic memory and the brain: structure and processes. Curr. Opin. Neurobiol. 11 (2), 194–201.

Mayes, A., Montaldi, D., Migo, E., 2007. Associative memory and the medial temporal lobes. Trends Cogn. Sci. 11 (3), 126–135.

McClelland, J.L., McNaughton, B.L., O'Reilly, R.C., 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev. 102 (3), 419–457.

McCloskey, M., Cohen, N.J., 1989. Catastrophic interference in connectionist networks: the sequential learning problem. In: Gordon, H.B. (Ed.), Psychology of Learning and Motivation, 24. Academic Press, pp. 109–165.

McQueen, J.M., Norris, D., Cutler, A., 1994. Competition in spoken word recognition: spotting words in other words. J. Exp. Psychol. Learn. Mem. Cogn. 20, 621–638.

Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., Rosenbaum, R.S., 2006. The cognitive neuroscience of remote episodic, semantic and spatial memory. Curr. Opin. Neurobiol. 16 (2), 179–190.

Nadel, L., Moscovitch, M., 1997. Memory consolidation, retrograde amnesia and the hippocampal complex. Curr. Opin. Neurobiol. 7 (2), 217–227.

Newman, S.D., Twieg, D., 2001. Differences in auditory processing of words and pseudowords: an fMRI study. Hum. Brain Mapp. 14 (1), 39–47.

Patterson, K., Nestor, P.J., Rogers, T.T., 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. Nat. Rev. Neurosci. 8 (12), 976–987.

Penny, W.D., Holmes, A.P., Friston, K.J., 2003. Random effects analysis. Human brain function. In: Frackowiak, R.S.J., Friston, K.J., Frith, C., et al. (Eds.), Academic Press.

Poser, B.A., Versluis, M.J., Hoogduin, J.M., Norris, D.G., 2006. BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. Magn. Reson. Med. 55 (6), 1227–1235.

Prabhakaran, R., Blumstein, S.E., Myers, E.B., Hutchison, E., Britton, B., 2006. An event-related fMRI investigation of phonological–lexical competition. Neuropsychologia 44 (12), 2209–2221.

Price, C.J., 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. Ann. N. Y. Acad. Sci. 1191, 62–88.

Price, C.J., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. NeuroImage 62 (2), 816–847.

Raettig, T., Kotz, S.A., 2008. Auditory processing of different types of pseudo-words: an event-related fMRI study. NeuroImage 39 (3), 1420–1428.

Scoville, W.B., Milner, B., 1957. Loss of recent memory after bilateral hippocampal lesions. J. Neurol. Neurosurg. Psychiatry 20 (1), 11–21.

Squire, L.R., Zola-Morgan, S., 1991. The medial temporal lobe memory system. Science 253 (5026), 1380–1386.

Suzuki, W.A., 2007. Integrating associative learning signals across the brain. Hippocampus 17 (9), 842–850.

Tamminen, J., Gaskell, M.G., 2008. Newly learned spoken words show long-term lexical competition effects. Q. J. Exp. Psychol. 61 (3), 361–371.

Tamminen, J., Payne, J.D., Stickgold, R., Wamsley, E.J., Gaskell, M.G., 2010. Sleep spindle activity is associated with the integration of new memories and existing knowledge. J. Neurosci. 30 (43), 14356–14360.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15 (1), 273–289.

Winocur, G., Moscovitch, M., 2011. Memory transformation and systems consolidation. J. Int. Neuropsychol. Soc. 17 (5), 766–780.