

Interaction between episodic and semantic memory networks in the acquisition and consolidation of novel spoken words



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

^a Radboud University, Behavioural Science Institute, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands

^b Radboud University, 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, P.O. Box 310, 6500 AH Nijmegen, The Netherlands

ARTICLE INFO

Article history:

Received 28 May 2015

Revised 8 April 2016

Accepted 23 May 2016

Available online 10 June 2016

Keywords:

Novel word learning

fMRI

Consolidation

Integration

Episodic memory

Semantic memory

Hippocampus

Posterior middle temporal gyrus

Angular gyrus

Inferior frontal gyrus

ABSTRACT

When a novel word is learned, its memory representation is thought to undergo a process of consolidation and integration. In this study, we tested whether the neural representations of novel words change as a function of consolidation by observing brain activation patterns just after learning and again after a delay of one week. Words learned with meanings were remembered better than those learned without meanings. Both episodic (hippocampus-dependent) and semantic (dependent on distributed neocortical areas) memory systems were utilised during recognition of the novel words. The extent to which the two systems were involved changed as a function of time and the amount of associated information, with more involvement of both systems for the meaningful words than for the form-only words after the one-week delay. These results suggest that the reason the meaningful words were remembered better is that their retrieval can benefit more from these two complementary memory systems.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

In everyday life, we regularly encounter novel words and new concepts, even as adults. How do we encode and consolidate novel words? The Complementary Learning Systems (CLS) account posits two memory networks that interact when we acquire new information (McClelland, McNaughton, & O'Reilly, 1995). The episodic memory network, with the hippocampus as the critical component, is important at the initial stage of learning when discrete episodes of experiences are encoded (e.g., an experience of hearing a novel word). After a subsequent period of consolidation, a shift towards more systematic, lexicalised coding of the memory representation in a distributed neocortical network occurs. The current study explores this consolidation process for words learned with and without meanings.

Through consolidation, novel words become integrated into the existing mental lexicon. It is after this integration process that interaction between novel words and existing words is observed

* Corresponding author at: Radboud University, Donders Institute, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands.

E-mail address: Atsuko.Takashima@donders.ru.nl (A. Takashima).

(Davis & Gaskell, 2009). Multiple studies have shown that this lexical integration effect can be measured behaviourally as a slowing down of processing of those existing words with which the novel words overlap in word-form (Bakker, Takashima, van Hell, Janzen, & McQueen, 2014; Davis, Di Betta, Macdonald, & Gaskell, 2009; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003; Henderson, Weighall, Brown, & Gaskell, 2013). In these studies, participants were taught novel words such as *cathedruke* (a word derived from the existing word *cathedral*) and then tested on the detection of a pause in the base words from which the novel words were derived (e.g., *cathe--dral*). It was hypothesised that once the novel words are integrated into the mental lexicon, they should interfere with the processing of phonologically neighbouring words. As a consequence, slowing down of responses in the pause detection task should be observed for the base words (*cathedral*) of the trained novel words (*cathedruke*). This phenomenon was indeed observed, but was found in several studies only after a delay of 24 h or a week (Bakker et al., 2014; Gaskell & Dumay, 2003). Sleep is thought to be one of the driving factors of this integration process (Dumay & Gaskell, 2007; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010), although some studies

suggest that sleep may not be necessary and that consolidation can be found directly after training (Kapnola, Packard, Gupta, & McMurray, 2015; Lindsay & Gaskell, 2013; Szmalec, Page, & Duyck, 2012). Sleep seems to be, in any case, beneficial for the new words to be more fully integrated into the lexico-semantic network (Dumay & Gaskell, 2007).

This lexical integration effect was originally tested for novel word forms learned without meanings. Recently, integration of novel words with meanings has also been observed, both at the word-form level (Takashima, Bakker, van Hell, Janzen, & McQueen, 2014) and at the word-meaning level (Clay, Bowers, Davis, & Hanley, 2007; Tamminen & Gaskell, 2013; van der Ven, Takashima, Segers, & Verhoeven, 2015). In previous functional Magnetic Resonance Imaging (fMRI) work, we demonstrated that novel words elicited more activation in the neocortex after a 24-h consolidation period, suggesting a neocortical retrieval route for words that had undergone a period of consolidation (Takashima et al., 2014). This was more pronounced for spoken novel words associated with novel visual pictorial information (unfamiliar objects) compared to words learned only as meaning-less phonological forms. In contrast, the behavioural lexical integration effect that emerged after 24 h was significant only for the form-only words. Why did the picture-associated novel words show less integration? We argued that the non-verbalisable nature of the pictures could have caused a delay in integration. In natural language learning, however, new words are normally accompanied by their corresponding conceptual meanings. In the present study, we aimed to find signatures of integration for phonological word forms associated with visual information by (1) testing at a delay of one week instead of 24 h, and (2) by making the visual referents easier to conceptualise (see Fig. 1).

One way to think of stabilisation of novel words from a systems-consolidation point of view is that with consolidation, the word becomes less dependent on episodic memory (a specific memory of the event in which a new word was acquired). Instead it becomes more conceptual in nature, that is, the word is no longer associated with the context in which it was experienced, but more with its concept or meaning. This semantic representation may be different from a memory representation of a word that comprises an experience. For example, one might associate the novel word *kathedrook* with a Japanese harp-like instrument upon seeing the object and hearing the word. Remembering the experience of seeing an actual *kathedrook* and hearing its name involves retrieving an episodic memory, whereas being able to describe what a *kathedrook* is would depend on processing in the semantic memory system. These two representations, even though they both refer to the object *kathedrook*, may be stored in different structures of the brain or may be accessed via different memory networks. The hippocampal episodic network is likely to be utilised when we remember the experience of seeing an object and hearing its name (Moscovitch, Nadel, Winocur, Gilboa, & Rosenbaum, 2006; Squire & Knowlton, 1994), whereas the semantic memory system, in a distributed neocortical network, is likely to be utilised when we retrieve the semantic information associated with the word (Binder & Desai, 2011; Martin & Chao, 2001; Patterson, Nestor, & Rogers, 2007; Price, 2010). After some time, we might not remember exactly where or when we encountered new words, but we nevertheless remember many of them and know what they mean. This is because the nature of the two memory systems differs: the episodic hippocampal memory is prone to decay over time, whereas the semantic memory is more stable over time.

The second aim of this study was to investigate the stability of novel semantic memory representations. We measured response latency on a task in which participants retrieved the matching referent when cued with a trained novel word, just after training (association memory test, Day1) and again after a week's delay

(Day8). If the referent is presented in a form that is not exactly the same as the one shown during the training, but still refers to the same concept, one cannot rely on episodic memory to perform this task. In order to choose the correct referent, one would instead have to rely on generalised semantic memory. We tested the accuracy and latency of retrieval by presenting referents in the same way as during training (Meaning-same condition) and in a different form (Meaning-similar condition). For example, if they were trained on the word *kathedrook* in association with a photo depicting the word, they will have the same photo that was shown during the training as the correct response option for the Meaning-same condition at test, whereas for the Meaning-similar condition, another photo of a *kathedrook* would be presented (see Fig. 1). We hypothesised that participants would be more accurate and faster in choosing the associated meaning information if associative memory strength was stronger. Referents that were the same as those presented during training could be chosen by accessing both the episodic memory representation and the lexical-semantic memory representation, whereas for physically different referents, access to a generalised semantic concept is necessary. Thus we assumed that reaction times (RTs) on the two association memory tests would be slower in the Meaning-similar condition than in the Meaning-same condition. After a delay, however, the Meaning-same condition may slow down due to decay of episodic memory traces, but this should not be the case in the Meaning-similar condition if semantic representations are stable over time.

In recent years, studies on the neural correlates of novel word representations have increased in numbers (e.g., Cornelissen et al., 2004; Davis et al., 2009; Mestres-Missé, Càmarà, Rodríguez-Fornells, Rotte, & Münte, 2008; Paulesu et al., 2009; Takashima et al., 2014). Semantic representation is thought to be distributed across a wide range of cortical areas (reviewed in: Binder & Desai, 2011; Martin, 2007; Patterson et al., 2007; Price, 2010), and many of these regions have been reported to be involved in the processing of novel words as well. Semantic information can be related to modality-specific (e.g. visual, auditory, tactile) and more abstract verbal information, where sensory specific information is thought to recruit sensory-specific cortical areas (reviewed in Martin, 2007; Patterson et al., 2007). With respect to the processing of multi-modal information, Damasio introduced Convergence zone theory, which states that sensory-specific information converges into the multi-modal units in association areas of the brain (Damasio, 1989). Building on this theory, Simmons and Barsalou (2003) proposed a hierarchical distribution model in which posterior regions are more modality specific, and anterior structures more modality general (further reviewed in Meyer & Damasio, 2009). Recent studies on the semantic system also suggest a role for posterior higher-order associative areas in the temporo-parietal junction including the angular gyrus, which acts as a convergence zone in a distributed semantic network (Binder, Desai, Graves, & Conant, 2009; Graves, Binder, Desai, Conant, & Seidenberg, 2010; Vandenberghe et al., 2013), and the posterior middle temporal gyrus, which acts as a lexical hub (Gow, 2012; Hickok & Poeppel, 2004, 2007). Concepts of novel words can be sensory-specific if the novel word is associated with input in a specific modality, such as visual or auditory information. It can also be more abstract in nature if the concept is conveyed verbally (e.g. reading the definition in a dictionary). Does the representation of a novel word differ according to the type of conceptual information that accompanies it? To our knowledge, no imaging study has systematically compared the impact of different types of conceptual information on novel word learning and consolidation. The third aim of the present study was therefore to investigate whether there are differences in the consolidation trajectory when novel word-forms were associated with verbal versus pictorial referents.

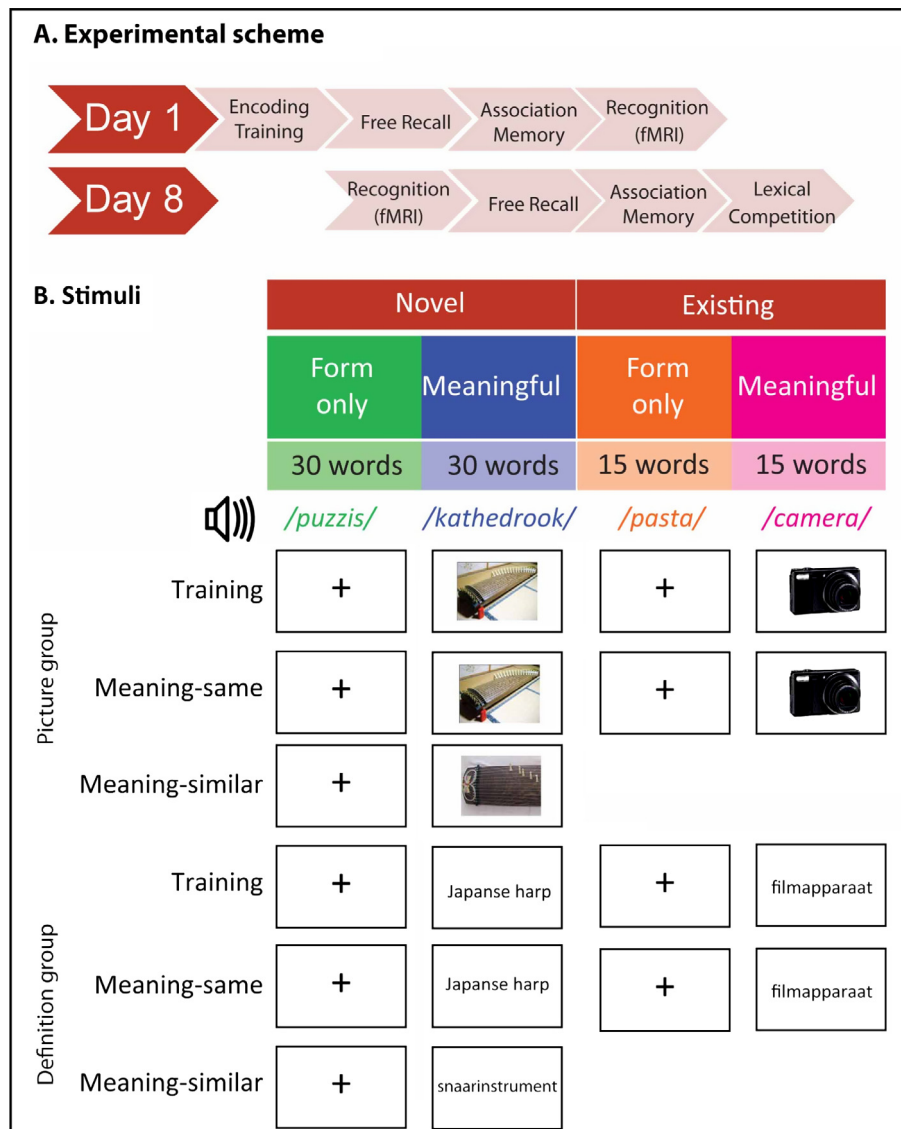


Fig. 1. Experimental setup. Testing took place on Day1 and Day8. On Day1 participants were first trained on a list of novel and existing words, half presented with and half without the words' meanings. The picture group received pictures as visual representations of the words' meanings, and the definition group received short definitions printed in Dutch. A series of memory tests followed. One week later, on Day8, participants completed a series of memory tests and a lexical competition task. The recognition memory test took place in an MRI scanner on Day1 and Day8 (A). Examples of stimuli in different conditions (B). N.B. *Japanese harp* = Japanese harp, *snarinstrument* = string instrument, *filmapparaat* = photo device.

To answer the above questions, participants were trained on a set of novel words, half with meanings (Meaningful words) and half without meanings (Form-only words). One group received the word meanings in the format of pictures of unfamiliar objects (Picture group); the other received the word meanings in the format of verbal definitions of unfamiliar objects (Definition group). Memory for the words was tested in the fMRI scanner on two occasions, once directly after training (Day1) and again a week later (Day8) to probe changes in brain areas involved in retrieval of the trained words (see Fig. 1). We measured the brain activation increase/decrease for successful recognition of the trained novel words relative to untrained words. We also investigated the brain response as a network of activation patterns using functional connectivity analysis. Since the words were cued auditorily, we investigated the brain areas that co-activated with the bilateral auditory cortex when hearing the words for recognition. In a previous study, we reported that lexical integration was observed not only as an increase in activity, but also an increase in functional connectivity

in the left posterior middle temporal gyrus (pMTG) (Takashima et al., 2014). Given that this area is known to integrate auditory, visual and semantic information (Gow, 2012; Hickok & Poeppel, 2004), we reasoned that once novel words became integrated into the lexical network, the left pMTG would start to act as a hub region to associate audio-visual aspects of the new words. To test this hypothesis we performed functional connectivity analysis with the seed region in the left pMTG. To probe lexical integration of the novel words behaviourally, we measured RTs during a lexical competition task (pause detection) on the base words that the novel words were derived from. To test the stability of episodic and semantic memory, we performed a cued-recall task using the trained words as cues and same/similar referents (pictures/definitions) as choice options on both testing days (Day1/Day8).

In summary, this study aimed to investigate the neural representations of novel words encoded with/without meanings, and asked whether there is a shift in memory representation with consolidation. In line with the CLS account, we hypothesised greater

involvement of the episodic hippocampal memory system at the initial stage of retrieval, but a shift towards lexical-semantic representation distributed across neocortical areas after consolidation, with a prominent role for the left pMTG as a hub region that links sound to meaning representations of the novel word. Additionally, we hypothesised that this shift would be more pronounced for novel words with meanings, and this would be reflected behaviourally by a higher degree of integration compared to form-only words in the lexical competition task. Furthermore, we expected memory decay for semanticised memories to be slower than for episodic memories in the association memory test. We also asked whether the consolidation trajectory is different (spatially or temporally) when participants were provided with pictorial material or with verbal definitions of novel words. If verbal and pictorial concept representations are distinct, we may find differences in brain activation patterns related to novel word retrieval, depending on whether the novel words were associated with a pictorial or a verbal referent at the time of encoding.

2. Material and methods

2.1. Participants

Fifty right-handed native speakers of Dutch (38 females, $M = 22.3$ years, range 18–32 years) participated in the experiment in return for course credit or monetary compensation. Participants had no history of neurological or language-related disorders, and reported having normal or corrected-to-normal vision and hearing. The participants were assigned to either the Picture group ($n = 24$) or the Definition group ($n = 26$). Four participants from the Definition group were excluded from the analyses (one male for participating in another similar study, one male for not completing the study, one female due to an asymptomatic mass in the prefrontal cortex distorting the functional brain images, and one female due to RTs longer than 10 s on more than 25% of the trials during the recognition test in the scanner), leaving 22 participants in the Definition group.

2.2. Stimulus materials

2.2.1. Novel spoken words

Word-form materials consisted of digital recordings of 39 disyllabic and 81 trisyllabic word pairs, spoken by a young female native speaker of Dutch, each including an existing monomorphemic Dutch noun (base word) and a pseudoword derived by substituting the final vowel 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). Subsets of these stimuli were used by Bakker et al. (2014) and Takashima et al. (2014). Base words were between 5 and 9 phonemes long ($M = 6.8$) and had a lemma frequency between 1 and 49 per million ($M = 8.9$) in CELEX (Baayen, Piepenbrock, & van Rijn, 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 position and was always located before the final vowel. An extra 60 Dutch words (hereafter labelled as existing words), 40 disyllabic and 20 trisyllabic words, between 5 and 10 phonemes long ($M = 6.2$) and with a lemma frequency between 1 and 49 per million ($M = 11.8$), were used in a control condition. The 120 pseudowords (hereafter labelled as novel words) were divided into four lists of 30 words each and were used in one of the four conditions (trained - Form only; trained - Meaningful; Untrained - Day1; Untrained - Day8). The sixty existing words were also divided into four lists of 15 words each, and were assigned to one of the four conditions. The allocation of lists to

conditions was counterbalanced across participants (see Fig. 1B for example stimuli).

2.2.2. Pictorial materials

Two colour images of the 60 existing words and two images of 30 Japanese objects not familiar to Dutch participants (confirmed by ten independent non-Asian participants rating the items as low in familiarity), were downloaded from freely available images on the internet (see Fig. 1B for an example). All images were in colour and 400×300 pixels in resolution. One image of each existing and novel word was assigned to one list and the other image was assigned to a second list. One list was used for training and as the "Meaning-same" condition in the association test, and the other list was used for the "Meaning-similar" condition in the association test. The assignment of the list to condition was counterbalanced between participants.

2.2.3. Definition

Definitions consisted of short descriptions ($M = 10.4$ words, range 3–20 words) of the objects included in the pictorial materials (see Fig. 1 for an example). Shorter versions (one version for existing words and two versions for novel words) were created from the longer version of the definitions ($M = 1.6$ words, range 1–4 words). Each of the two short versions of the novel words were assigned to one of the two lists. One list was used for training and as the "Meaning-same" condition in the association test, and the other list was used for the "Meaning-similar" condition in the association test. The assignment of the list to condition was counterbalanced between participants.

2.3. Procedure

On Day1, the participants were exposed to a list of novel and existing spoken words (encoding phase). The memory for these words was further strengthened through repeated exposure (training phase). This was followed by a free recall memory test, association memory test, and finally a recognition memory test in the MRI scanner. One week later, the participants returned to the laboratory (Day8). First their recognition memory for the trained words was tested in the scanner, followed by free recall and association tests outside of the scanner. Finally a lexical competition task (pause detection) was administered (see Fig. 1A). All tasks were presented using the software Presentation (www.neurobs.com).

2.4. Tasks

2.4.1. Encoding and training

The training session consisted of two sub-parts: an encoding phase and a training phase. In the encoding phase, participants were exposed to all to-be-trained novel (30 Meaningful, 30 Form-only) and existing words (15 accompanied by a relevant picture/definition and 15 Form-only), one at a time. For words in the Meaningful condition, the associated picture or definition was presented simultaneously with the auditory word presentation, and for words in the Form-only condition, a fixation cross appeared, and stayed on the screen until the participant pressed the button for the next trial. This was followed by a training session consisting of eight blocks. For each block, all words (novel and existing) were presented once in a randomised order, together with four response options of which one was always correct: a fixation cross, two pictures/definitions associated with one of the trained novel words, and one picture/definition associated with one of the trained existing words. Participants were instructed to select the corresponding picture or definition. They were asked to choose the fixation cross option for words that they thought had no meaning (e.g. the words

in the Form-only condition), or if they thought that none of the picture/definition options matched the word. Feedback about whether the response was correct or incorrect was given on each trial, and at the end of each block feedback on the number of correct responses was given. For blocks one to six, the word was presented together with the correct picture/definition (or a fixation cross in case of Form-only words) after each trial, and for the last two blocks, the word together with the correct picture/definition/fixation cross was repeated only when the participant gave an incorrect response.

2.4.2. Free recall

Participants were given three minutes to recall and produce as many of the trained words as possible. The spoken responses were recorded for further analysis.

2.4.3. Association memory test

This test was similar to the training session. All novel and existing words from the encoding/training session were presented on the screen once in randomised order, together with four response options. As in the training session, the participants' task was to choose the corresponding picture/definition/fixation cross. Participants were instructed to press the button for the corresponding option. As in the training session, they were asked to choose the fixation cross if they thought either that a word had no meaning (e.g. a Form-only word), or that none of the picture/definition options matched the word. Different from the training session was that half of the pictures/definitions were the same as in the training session (Meaning-same condition), whereas the other half were slightly different from the ones used during the training session but still referred to the same object (Meaning-similar condition). In order to probe word-meaning retrieval through the semantic memory network, participants were informed that some pictures/definitions were not exactly the same as the ones presented during the training session, but that they should choose the similar one if the meaning matched the word (see Fig. 1B for an example). No feedback was given. RT was calculated as the time between onset of the word and button response. Assignment of the two picture/definition lists to either the Meaning-same or Meaning-similar condition was counterbalanced across participants and days, such that the words presented in the Meaning-same condition on Day1 were presented in the Meaning-similar condition on Day8, and vice versa.

2.4.4. Recognition test (fMRI)

In the scanner, participants heard the 60 novel words (30 Meaningful, 30 Form-only) and 30 existing words (15 Meaningful and 15 Form-only) that they had been trained on, intermixed with 30 untrained novel words and 15 untrained existing words. Words were presented through in-ear headphones and the volume was adjusted for each participant such that the words were clearly audible above the scanner noise. After a jittered inter-trial interval (ITI) of 2–6 s, a fixation cross on the screen turned from white to blue, signalling 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 coloured squares aligned horizontally, representing a scale from 'definitely old' to 'definitely new' by moving the cursor to the left or right by button presses using the left index and ring fingers, and then pressing a confirmation button using the left middle finger. No feedback was given. Untrained novel and untrained existing words were different for Day1 and Day8.

2.4.5. 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 & Dumay, 2003, Experiment 3). Base words of the 60 trained novel words (30 base words in the Meaningful condition, 30 base words in the Form-only condition), and 60 base words from the untrained novel word condition (Day1 and Day8 pooled) were presented, intermixed with 46 filler words 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 syllable. Pauses were always placed at zero-crossings using the Praat speech editor (Boersma & 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. RT was calculated as the time between pause onset and button response. For words without a pause, RTs were measured from the location of the pause onset in the pause-present version of the same word.

2.5. MRI data 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): 2.32 s, echo time: TE1 9.4 ms, TE2 21 ms, TE3 33 ms, TE4 44 ms, TE5 56 ms, 37 slices, ascending slice order, 3.0 mm slice thickness, 0.5 mm slice gap, inplane matrix size: 64 × 64, inplane field of view (FOV): 212 × 212 mm, flip angle: 90°. Slices were angulated in an oblique axial manner to reach whole-brain coverage (except for a part of the parietal cortex and the cerebellum). 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.

2.6. fMRI analyses

fMRI preprocessing and the creation of single subject general linear models (GLM) were performed in a similar way as reported in Takashima et al. (2014) using standard fMRI preprocessing tools in SPM8 (www.fil.ion.ucl.ac.uk) after the multiple echoes of the functional time series were first combined to one volume per time point (Poser, Versluis, Hoogduin, & Norris, 2006). In short, the functional images were realigned to the first volume, co-registered to the participant's anatomical scan, and slice-time corrected. Each participant's anatomical scan was segmented, and the normalisation parameter to the template MNI brain was estimated. This normalisation parameter was applied to each of the functional images, and lastly spatial smoothing with 8 mm full-width at half maximum Gaussian kernel was applied to all functional images.

For each participant, a GLM was applied to the functional images with the following conditions: trained Form-only novel words correctly recognised, trained Meaningful novel words correctly recognised, trained existing words correctly recognised, untrained novel words correctly rejected, untrained existing words correctly rejected, and trials of no interest (TNI) comprising all incorrect trials and trials on which the participant took longer than

10 s to respond. For all conditions except TNI, parametric modulators indicating the participant's level of confidence in their recognition response (1 = unsure, 2 = sure, 3 = very sure) were included as regressors of interest. These explanatory variables for the two sessions 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 six head motion regressors (three translations, three rotations) for each of the sessions. 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, contrast parameter images of the difference between conditions of interest were generated for each participant and then subjected to a second-level analysis (Penny, Holmes, & Friston, 2003), treating subjects as a random variable.

Psycho-physiological interaction (PPI) analysis implemented in SPM8 was performed to observe functional connectivity. PPI shows areas in which activity fluctuation correlates more strongly with activity fluctuation of the seed region of interest (ROI) (the physiological factor) in one condition over the other (the psychological factor) within a single run. We chose the bilateral auditory cortex as the first seed region since the memory for words was probed by auditory stimulation. We defined the bilateral auditory cortex by selecting the peak voxel in the superior temporal lobe for each hemisphere from group level activation in response to stimulus presentation, and took voxels within a sphere of 10 mm radius of the peak voxel. To probe memory-related connectivity differences, PPI effects of the Trained (Form-only + Meaningful) relative to Untrained novel words for each session were generated for each participant, and the contrast image reflecting this effect was tested on the second-level, treating each participant's contrast image as a random variable. To test whether the connectivity pattern changes when the novel word is accompanied by a meaning either in the form of a picture or a definition, as compared to when the novel word is only learnt with its word-form, we also tested the connectivity difference between Form-only and Meaningful conditions for the two sessions. For within-session contrasts, one-sample *t*-tests were performed. To test for changes in connectivity as a function of time, paired *t*-tests were performed.

Following up on the previous study where we showed that engagement of the left pMTG reflected the lexicalisation of the novel words (Takashima et al., 2014), we used the left pMTG as the seed region for the second functional connectivity analysis. The left pMTG was defined as a sphere with a 10 mm radius around the peak reported in the previous study [−54 −62 4] (Takashima et al., 2014). Since the meaningful novel words used in the current study were more concrete than those in Takashima et al. (2014), and there was more time for the words to consolidate, we expected a more robust connectivity pattern for the Meaningful words compared to Form-only words in the current study, especially on Day8. A contrast image of the PPI effect for the Meaningful relative to the Form-only condition was generated for each participant per session, and was tested at the group level using one-sample *t*-tests for each session.

Results of all second-level analyses, both activation and functional connectivity, were initially thresholded at the voxel-level, uncorrected $p < 0.001$ (cluster defining threshold). For the whole-brain search, cluster-size statistics were used as the test statistics, applying a cluster-extent threshold of family-wise error (FWE) correction of $p < 0.05$ (Hayasaka & Nichols, 2003), unless otherwise stated. We had *a priori* hypotheses about possible changes in the engagement of the hippocampus between unconsolidated and consolidated memory retrieval. To specifically address if disengagement of the hippocampus with consolidation could be observed, we performed ROI analyses on the bilateral hippocampi using an anatomically defined mask based on the AAL template (Tzourio-

Mazoyer et al., 2002) provided by WFU PickAtlas software (Maldjian, Laurienti, Kraft, & Burdette, 2003).

3. Results

3.1. Behavioural results

3.1.1. Recognition test

Recognition accuracy (proportion hits - false alarms) was high on Day1 and Day8 (all above 74%; see Table 1). A mixed-effects ANOVA on accuracy scores with two within-subjects factors Day (Day1, Day8) and Condition (Meaningful, Form-only) and one between-subjects factor (Picture, Definition group) revealed a main effect of Group ($F(1,44) = 6.97$, $p = 0.011$, $\eta_p^2 = 0.137$; $M_{Definition} = 88.7%$, $M_{Picture} = 83.5%$), Day ($F(1,44) = 14.91$, $p < 0.001$, $\eta_p^2 = 0.253$, $M_{Day1} = 88.9%$, $M_{Day8} = 83.3%$), and Condition ($F(1,44) = 48.63$, $p < 0.001$, $\eta_p^2 = 0.525$; $M_{Meaningful} = 90.6%$, $M_{Form-only} = 81.6%$) as well as Group \times Condition ($F(1,44) = 4.942$, $p = 0.031$, $\eta_p^2 = 0.101$) and Day \times Condition ($F(1,44) = 7.621$, $p = 0.008$, $\eta_p^2 = 0.148$) interaction effects. No three-way interaction ($F(1,44) = 0.654$, $p = 0.885$, $\eta_p^2 = 0.001$) and no Group \times Day interaction ($F(1,44) = 2.321$, $p = 0.135$, $\eta_p^2 = 0.050$) were observed.

Post-hoc *t*-tests revealed that the interaction effect of Group \times Condition was driven by the significantly better performance of the Definition group on Form-only words ($p = 0.002$; $M_{Picture} = 77.6%$; $M_{Definition} = 85.7%$), whereas the groups did not differ significantly in the Meaningful condition ($p = 0.161$; $M_{Picture} = 89.4%$, $M_{Definition} = 91.8%$). The interaction between Day and Condition suggests that Meaningful words were better retained across days than Form-only words (amount of forgetting calculated as the difference in accuracy scores Day1–Day8: $M_{Meaningful} = 3.2%$, $M_{Form-only} = 7.7%$, $p < 0.007$).

3.1.2. Free recall

A mixed-effects ANOVA on the number of words recalled during the Free recall sessions (see Table 1) with two within-subjects factors Day (Day1, Day8) and Condition (Meaningful, Form-only) and one between-subjects factor Group (Picture, Definition) revealed a main effect of Day ($F(1,44) = 32.04$, $p < 0.001$, $\eta_p^2 = 0.421$, $M_{Day1} = 1.7$ words, $M_{Day8} = 2.8$ words) and of Condition ($F(1,44) = 61.71$, $p < 0.001$, $\eta_p^2 = 0.584$, $M_{Meaningful} = 3.5$, $M_{Form-only} = 1.1$) as well as a Day \times Condition ($F(1,44) = 11.07$, $p = 0.002$, $\eta_p^2 = 0.201$) interaction. The number of words recalled for the Meaningful condition increased over time (post-hoc *t*-test comparing Day1 and Day8: $p < 0.001$), whereas Form-only words did not ($p = 0.078$). No main effect of Group ($F(1,44) = 0.021$, $p = 0.887$, $\eta_p^2 = 0.000$; $M_{Picture} = 2.3$, $M_{Definition} = 2.3$) or interactions of Group \times Day ($F(1,44) = 1.17$, $p = 0.286$, $\eta_p^2 = 0.026$), Group \times Condition ($F(1,44) = 2.96$, $p = 0.093$, $\eta_p^2 = 0.063$), nor Group \times Day \times Condition ($F(1,44) = 0.042$, $p = 0.839$, $\eta_p^2 = 0.001$) were found.

3.1.3. Association memory test

3.1.3.1. Accuracy

Accuracy scores were calculated as the proportion of correct responses (see Table 1). During the test, participants were instructed to choose the fixation cross option for words that they thought had no meaning (e.g. the words in the Form-only condition), or if they thought that none of the picture/definition options matched the word. An accurate meaning-related memory was hence required to choose the correct option for the Meaningful conditions, but not for the Form-only condition. If participants were unsure of their memory for the novel words, they may have defaulted to respond with the fixation cross option for words with weak or no memory for associated information. Since this is the correct option for the Form-only condition, this could lead to an inflation of correct response rates for the Form-only condition.

Table 1
Memory scores for recognition, free recall and association tests. Standard error of means (SE) in parentheses.

Group		Recognition performance				Free recall				Association test					
		Proportion hit - false alarms				Number of words recalled				% Correct response					
		mean% (SE)				mean (SE)				Form-only		Meaningful same		Meaningful similar	
		Form-only		Meaningful		Form-only		Meaningful		Form-only		Meaningful same		Meaningful similar	
Day1		Day8		Day1		Day8		Day1		Day8		Day1		Day8	
Picture	Performance	80.5 (2.5)	74.7 (2.7)	89.9 (1.6)	88.9 (1.8)	0.7 (0.1)	0.8 (0.2)	2.9 (0.3)	4.6 (0.4)	100 (0.1)	97 (0.9)	97 (0.9)	94 (1.9)	94 (1.5)	89 (2.5)
	Reaction time [ms]									1922 (90)	2325 (144)	1662 (68)	1745 (72)	1972 (85)	2036 (108)
Definition	Performance	90.6 (1.6)	80.7 (2.3)	94.6 (1.0)	89.0 (2.0)	1.1 (0.2)	1.7 (0.2)	2.3 (0.3)	4.3 (0.5)	99 (1.4)	87 (2.9)	95 (1.4)	81 (3.9)	85 (3.0)	62 (4.1)
	Reaction time [ms]									2061 (187)	3416 (265)	2313 (133)	2724 (123)	3136 (155)	3258 (166)

We therefore investigated the type of incorrect responses made to the three different conditions. Most incorrect responses for the Meaning-same and Meaning-similar conditions were fixation-cross responses rather than different picture/definition responses (Picture group: Day1 97% of incorrect responses were fixation-cross responses, Day8 92%; Definition group: Day1 91%, Day8 74%). Because of this confound, we did not analyse the accuracy scores for the association memory test further.

3.1.3.2. Reaction time

For each participant, incorrect trials and RTs below 100 ms were discarded (0% of the trials were responded below 100 ms). Furthermore, correct trials with RTs 2 SD above or below the mean were treated as outliers and were omitted from the analysis (Day1 6.9%, Day8 12.7%). Since the Picture and Definition groups differed in the visual inputs of the response options, and reading definitions would inevitably take longer than seeing objects on the screen, we conducted separate repeated-measures ANOVAs for each group with two within-subjects factors Day (Day1, Day8) and Condition (Meaning-same, Meaning-similar, Form-only).

For the Picture group, the main effects showed that responses were slower on Day8 than Day1 ($F(1,23) = 14.946$, $p = 0.011$, $\eta_p^2 = 0.247$; $M_{Day1} = 1852$ ms, $M_{Day8} = 2035$ ms) and the Meaningful-same condition was the fastest and the Form-only condition slowest ($F(1.6,36.4) = 31.189$, $p < 0.001$, $\eta_p^2 = 0.576$; $M_{Meaning-same} = 1703$ ms, $M_{Meaning-similar} = 2004$ ms, $M_{Form-only} = 2123$ ms). The Day \times Condition interaction effect was also significant ($F(1.9,44.8) = 6.996$, $p = 0.002$, $\eta_p^2 = 0.233$). Slowing down occurred most for Form-only condition ($\Delta Day8 - Day1 = 403$ ms, $p < 0.001$), followed by the Meaning-same condition ($\Delta Day8 - Day1 = 84$ ms, $p = 0.162$), and least for the Meaning-similar condition ($\Delta Day8 - Day1 = 64$ ms, $p = 0.540$).

Likewise, for the Definition group, the main effects showed that responses were slower on Day8 than Day1 ($F(1,21) = 7.544$, $p = 0.001$, $\eta_p^2 = 0.416$; $M_{Day1} = 2724$ ms, $M_{Day8} = 2991$ ms) and Meaningful-same condition was the fastest and the Meaning-similar condition slowest ($F(1.4,28.9) = 10.091$, $p = 0.002$, $\eta_p^2 = 0.325$; $M_{Meaning-same} = 2518$ ms, $M_{Form-only} = 2738$ ms, $M_{Meaning-similar} = 3197$ ms). The significant Day \times Condition interaction effect ($F(1.8,38.1) = 26.679$, $p < 0.001$, $\eta_p^2 = 0.560$) was driven by the most slowing down observed for the Form-only condition ($\Delta Day8 - Day1 = 1354$ ms, $p < 0.001$), followed by the Meaning-same condition ($\Delta Day8 - Day1 = 411$ ms, $p = 0.021$). Meaning-similar condition, on the other hand, did not slow down at the delayed test ($\Delta Day8 - Day1 = 122$ ms, $p = 0.576$).

3.1.4. Lexical competition task (pause detection)

Incorrect trials, trials with no response and RTs below 100 ms were discarded. For each participant, trials with RTs that were 2

SD above or below the mean, calculated separately for pause-present and pause-absent conditions, were omitted. This resulted in discarding of 4.5% of the pause-present and 4.3% of the pause-absent trials.

A mixed-effect repeated-measures ANOVA with two within-subjects factors Pause (present, absent), Condition (Meaningful, Form-only, Untrained), and one between-group factor (Picture, Definition) was performed. The main effect of Pause showed that pause-absent items were responded to faster than pause-present items ($F(1,44) = 15.11$, $p < 0.001$, $\eta_p^2 = 0.256$; $M_{pause-present} = 684$ ms, $M_{pause-absent} = 645$). The main effect of Condition was significant ($F(1.96,86.29) = 3.32$, $p = 0.042$, $\eta_p^2 = 0.070$, $M_{Untrained} = 656$ ms, $M_{Form-only} = 664$ ms, $M_{Meaningful} = 672$ ms): responses to base-words in the Untrained condition were significantly faster than in the Meaningful condition (one-sided paired t -test $p < 0.008$) and marginally faster than in the Form-only condition (one-sided paired t -test $p < 0.054$), but responses were not different in the Form-only and Meaningful conditions (one-sided paired t -test $p = 0.133$). No main effect ($F(1,44) = 0.445$, $p = 0.508$, $\eta_p^2 = 0.010$) or any interactions with the factor group were observed, and no interactions with the Pause factor were observed (interactions: Group \times Pause $F(1,44) = 2.70$, $p = 0.108$, $\eta_p^2 = 0.508$; Group \times Condition $F(1,44) = 0.089$, $p = 0.912$, $\eta_p^2 = 0.002$; Pause \times Condition $F(1.82,80.02) = 0.400$, $p = 0.652$, $\eta_p^2 = 0.009$; three-way $F(1.82,80.02) = 0.284$, $p = 0.733$, $\eta_p^2 = 0.006$).

3.1.5. Summary – behavioural results

In summary, more words were recognised (in the recognition test in the scanner) and recalled (in the free recall test) when they had been learned together with their meanings. The modality in which the concept was acquired (picture/definition) also had an effect: the recognition score was better in the Definition group, but this was mainly driven by better performance on the Form-only words for this group; meaningful words did not differ between the two groups. Free recall performance did not differ between the two groups. The speed with which associated information was correctly retrieved slowed down for both the Form-only and Meaning-same conditions, whereas the Meaning-similar condition did not slow down with time. In accordance with our prediction, we observed a lexical competition effect on Day8 (i.e., slowing down in detecting pauses in the phonologically neighbouring words of the trained novel words) for both groups.

3.2. Imaging results

3.2.1. Activation

In order to probe memory representations of the novel words, and their changes with consolidation, we first compared the activation patterns for correctly recognised trained items relative to

correctly rejected untrained novel items. More specifically, difference contrast images between trained and untrained novel words were generated separately for Form-only words and Meaningful words for each session per participant, resulting in the following four contrasts: Form-only > Untrained Day1, Meaningful > Untrained Day1, Form-only > Untrained Day8, and Meaningful > Untrained Day8. These four contrast images for every participant were investigated on the second level using a full-factorial design with three factors: Group (Picture, Definition), Day (Day1, Day8) and Condition (Form-only, Meaningful). This revealed a main effect of Day and of Condition, as well as two-way interactions of Group \times Condition and Day \times Condition (see Table 2). No significant clusters were observed for the three-way interaction (all clusters above $p_{FWE} \geq 0.489$), the two-way interaction of Group \times Day (all $p_{FWE} \geq 0.774$), or the main effect of Group.

To further investigate the Group \times Condition interaction effect, we looked into the contrast Form-only versus Meaningful trained

words for the two days collapsed. In order to examine the difference in activity patterns for the two conditions, a contrast image of Form-only (Day1 + Day8) versus Meaningful (Day1 + Day8) was generated for each participant, and using a one-sample t -test, the activation difference was tested for each group separately. For the Picture group, no areas showed an activation pattern that was greater for the Form-only relative to Meaningful condition. The reverse contrast (Meaningful > Form-only) showed a distributed pattern, including the left hippocampus when small volume correction (SVC) on the anatomical hippocampus mask was applied (Table 3). For the Definition group, even though there was more activation for the Meaningful condition overall, no clusters survived the statistical threshold (all $p \geq 0.108$). Thus, the interaction effect was mainly driven by the greater activation for Meaningful as compared to Form-only words in the Picture group.

Further investigation on the Day \times Condition interaction effect was performed by contrasting Form-only and Meaningful

Table 2
Main effects of brain activation differences.

Factors	Cluster		Peak voxel				Anatomical region	
	P_{cluster}	Size	z-score	x	y	z		
Group	No significant clusters (all $p_{FWE} \geq 0.774$)							
Day	Day1 > Day8	<0.001	523	4.73	-6	-66	34	L precuneus
		0.037	126	3.89	-10	-50	34	L posterior cingulate
				3.98	-40	-64	38	L angular
	Day1 < Day8	No significant clusters						
Condition	Form-only > Meaningful	No significant clusters						
	Meaningful > Form-only	<0.001	1833	6.21	-30	-36	-22	L fusiform
				3.85	-18	-16	-18	L hippocampus
				3.82	-50	-50	-16	L inferior temporal
				3.81	-18	-4	-26	L parahippocampal
				3.79	-28	6	-2	L putamen
		<0.001	3372	6.1	-8	56	4	L medial superior frontal
				5.14	-10	46	-10	L medial orbital frontal
				5.08	-8	36	-4	L anterior cingulate
				4.58	-12	38	-14	L rectus
				4.58	-48	34	6	L inferior frontal pars triangularis
				4.2	-18	28	50	L superior frontal
				4.15	8	42	-2	R medial orbital frontal
				4.09	-24	30	-12	L inferior orbital frontal
				4.04	-10	46	36	L medial superior frontal
				4.04	-20	32	44	L superior frontal
				3.97	-40	24	-20	L superior temporal pole
		<0.001	2043	5	-6	-58	24	L precuneus
				4.53	-6	-44	34	L posterior cingulate
				4.47	-2	-38	36	L middle cingulate
				4.24	-4	-66	28	L cuneus
				4.22	-14	-46	6	L calcarine
				4.19	-6	-56	4	L lingual
		0.02	242	4.83	-14	-16	10	L thalamus
		<0.001	1115	4.8	44	-16	48	R precentral
				4.01	42	-22	46	R postcentral
				3.82	14	-22	46	R middle cingulate
				3.45	8	-18	50	R supplementary motor area
		<0.001	833	4.76	32	2	12	R putamen
				4.3	24	-2	4	R pallidum
		0.005	323	4.49	-8	4	34	L middle cingulate
				3.57	4	-2	38	R middle cingulate
		<0.001	834	4.41	-46	-68	24	L middle temporal
				4.39	-32	-68	44	L inferior parietal
		0.032	216	4.37	60	-8	-22	R middle temporal
				4.23	52	8	-30	R middle temporal pole
		0.036	209	4.25	36	30	-18	R orbital inferior frontal
				3.85	44	22	-30	R superior temporal pole

L: left, R: right, FWE: family-wise error corrected.

Table 3
One-sample *t*-test Form-only versus Meaningful (Day1 + Day8).

	Cluster		Voxel z-score	MNI coordinates			Anatomical region
	<i>p</i> FWE	Size		x	y	z	
Picture group							
<i>Meaningful > Form-only</i>							
<0.001	1541	5.22	-14	44	-6	L orbital medial frontal	
			-8	36	0	L anterior cingulate	
			-6	58	2	L medial superior frontal	
<0.001	1436	4.61	-12	-42	42	L middle cingulum	
			-2	-58	24	L precuneus	
			4	-52	8	R lingual	
			3.94	-10	-58	L precuneus	
			3.78	8	-56	R calcarine	
			0.001	396	4.47	-50	-72
0.015	223	4.41	-40	-82	30	L middle occipital	
			-60	-12	-20	L middle temporal	
<0.001	637	3.88	-62	-22	-20	L inferior temporal	
			4.38	-30	-32	-18	L fusiform (ATL)
0.017	217	4.33	-28	-28	-16	L parahippocampal	
			4.3	-30	10	-38	L middle temporal pole
			3.49	-46	-22	-24	L inferior temporal
0.047	169	3.66	-12	54	24	L superior frontal	
			-8	50	18	L medial superior frontal	
0.008	258	4.31	62	-2	-24	R middle temporal anterior	
			4.26	4	-4	40	R middle cingulum
<0.001	450	3.25	-10	2	38	L middle cingulum	
			4.22	34	-18	62	R precentral
0.032 ^a	37	3.38	4.13	36	-20	48	R postcentral
			-30	-24	-14	L hippocampus	
<i>Form-only > Meaningful</i>							
No significant clusters							
Definition group							
<i>Meaningful > Form-only</i>							
No significant clusters							
<i>Form-only > Meaningful</i>							
No significant clusters							

^a Small volume correction using the hippocampal mask.

conditions separately for Day1 and Day8. Since there was no significant main effect of Group or a three-way interaction effect of Group × Day × Condition, we pooled the data of the two groups.

To see if the interaction was driven by the differences between conditions across days, difference contrasts of Form-only versus Meaningful, separately for Day1 and Day8 were generated for each participant and one-sample *t*-tests for Day1 and Day8 were performed separately. Day1 results revealed no significant difference between the Form-only and Meaningful conditions. Day8 results, on the other hand, revealed no clusters above threshold that were greater for the Form-only condition compared to the Meaningful condition, whereas multiple areas showed an increased activation pattern for the Meaningful condition relative to the Form-only condition (see Table 4).

To see if the interaction effect was affected by the difference within the conditions across days, we generated a contrast image of Day1 versus Day8, separately for the Form-only and Meaningful conditions. One-sample *t*-tests on these contrast images revealed that for the Meaningful condition, the engagement of the left precuneus decreased whereas activity in the pars triangularis of the inferior frontal gyrus increased with consolidation. The Form-only condition, on the other hand, showed a decrease in higher-order visual processing areas in the occipital cortex extending to the right parietal and precentral gyri over time, but no areas showed an increase in activation level over time. Against our prediction, we did not find that the hippocampus disengaged over time for both conditions, even when we

applied an anatomical hippocampal mask and restricted the search area to that ROI.

In sum, we did not observe a major difference between the Picture and the Definition groups. The interaction effect of Group × Condition was mainly driven by greater activation for Meaningful relative to Form-only words in the Picture group (Table 3). The interaction effect of Day by Condition was due to the Form-only condition showing a decrease in activation over time in the precuneus and right parietal cortex (Fig. 2B), whereas for the Meaningful condition, only the left precuneus decreased with time, and, notably, the left inferior frontal gyrus showed an increase over time (Fig. 2A). Furthermore, the difference between the Form-only and Meaningful conditions was not significant on Day1, but on Day8 distributed areas were more active for the Meaningful condition, including the left inferior frontal gyrus, left posterior middle temporal gyrus, and right hippocampus (Table 4, Fig. 2A).

3.2.2. Functional connectivity

3.2.2.1. Seed from the auditory cortex

To investigate which areas are functionally connected during successful recognition of the trained novel words, we applied a psycho-physiological interaction (PPI) analysis implemented in SPM8. PPI shows areas whose activity fluctuation correlates more strongly with activity fluctuation in the seed ROI for one condition over the other within a single run, which is a signature of functional connectivity that is greater during the processing of

Table 4

One-sample *t*-tests Form-only versus Meaningful conditions (both groups collapsed) separately for Day1 and Day8, and one-sample *t*-tests Day1 versus Day8 separately for Form-only and Meaningful conditions.

Day 1							
<i>Meaningful > Form-only</i>							
No significant clusters							
<i>Form-only > Meaningful</i>							
No significant clusters							
Day 8							
<i>Form-only > Meaningful</i>							
No significant clusters							
<i>Meaningful > Form-only</i>							
Cluster	Size	Voxel	MNI coordinates (peak)				
<i>p</i> FWE		<i>z</i> -score	x	y	z		
<0.001	3876	5.93	36	-22	48		R postcentral
		5.56	42	-16	48		R precentral
		5.01	-8	4	34		L middle cingulum
		4.19	12	-20	50		R supplementary motor area
		4.08	-6	-58	24		L precuneus
<0.001	3603	5.93	-8	60	6		L medial superior frontal
		4.57	-12	52	-8		L orbital medial frontal
		4.54	-18	52	18		L superior frontal
		4.5	8	44	-4		R orbital medial frontal
		4.24	-8	38	-6		L anterior cingulum
<0.001	962	5.73	-30	-36	-22		L fusiform
		3.78	-18	0	-4		L pallidum
		3.76	-24	-6	-16		L amygdala
		3.36	-46	-6	-26		L inferior temporal
0.004	322	4.82	-44	32	0		L inferior frontal pars triangularis
<0.001	1418	4.78	-48	-66	26		L angular
		4.51	-50	-48	26		L supramarginal
		4.04	-32	-72	34		L middle occipital
		3.79	-56	-60	16		L middle temporal
		3.38	-50	-54	38		L inferior parietal
<0.001	741	4.69	24	-6	2		R pallidum
		4.51	30	2	8		R putamen
		3.17	22	-12	-18		R hippocampus
0.029	210	4.2	-40	-60	-8		L inferior occipital
		4.08	-40	-54	-14		L fusiform
		3.91	-44	-58	-8		L inferior temporal
0.008	283	4.16	32	-26	-18		R parahippocampal
		3.91	36	-28	-18		R fusiform
		3.28	28	-16	-20		R hippocampus
Meaningful							
<i>Day1 > 8</i>							
	0.028	209	4.46	-8	-72	40	L precuneus
<i>Day8 > 1</i>							
	0.023	218	4.22	-44	28	6	L inferior frontal pars triangularis
Form-only							
<i>Day1 > 8</i>							
<0.001	876	4.69	-6	-72	38		L precuneus
		4.34	-16	-80	14		L calcarine
		3.97	-8	-68	26		L cuneus
<0.001	1062	4.68	44	-28	44		R postcentral
		3.87	30	-62	44		R angular
		3.81	38	-14	42		R precentral
		3.67	46	-42	40		R inferior parietal
		3.65	32	-68	42		R superior occipital
		3.49	36	-64	34		R middle occipital
		3.44	46	-36	42		R supramarginal
		0.044	180	3.95	-32	-56	42
		3.81	-30	-60	38	L middle occipital	
<i>Day8 > 1</i>							
No significant clusters							

FWE: family-wise error corrected, MNI: Montreal Neurological Institute, L: left, R: right.

words in one condition relative to the other. Since all trials were cued with auditorily presented words, we used the auditory

cortex as the first seed ROI to see which areas became co-active when novel words are perceived and recognised. We hypothesised

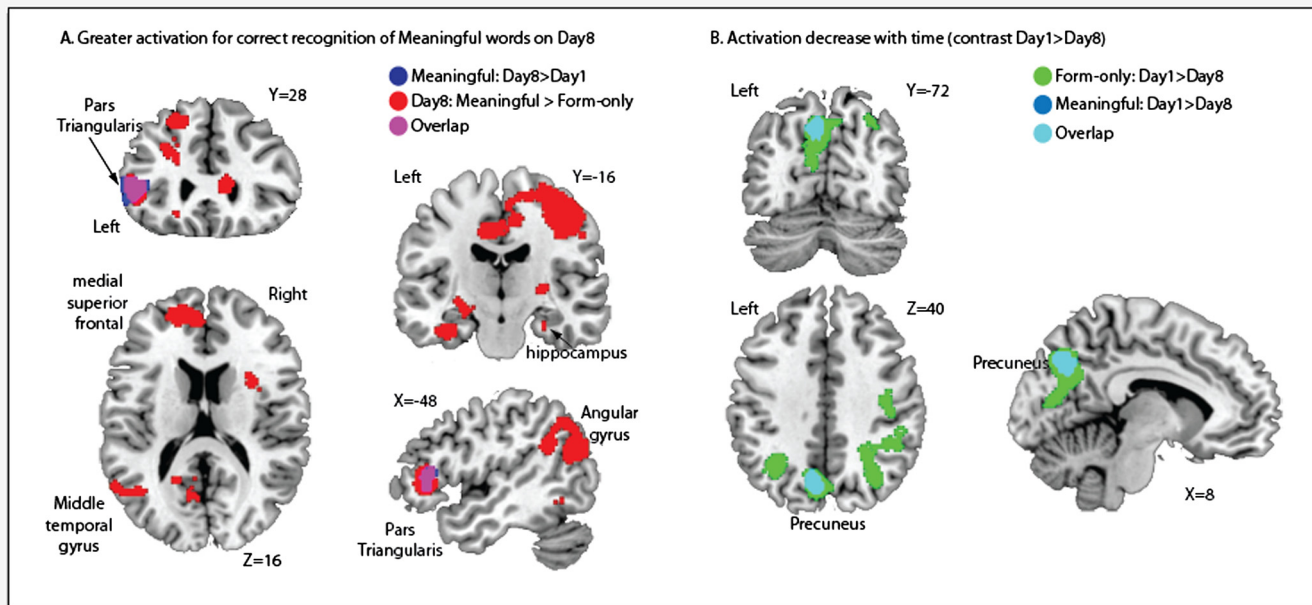


Fig. 2. Activation difference. (A) Greater activation for Meaningful words on Day8 compared with Meaningful words on Day1 (blue) and Form-only words on Day8 (red). Pink area shows overlapping areas for these two contrasts. (B) Activation that decreased with time for Form-only words (green) and Meaningful words (Blue). Light blue shows the overlapping areas for the two contrasts. All significant clusters (cluster level significant at $p_{FWE} < 0.05$) superimposed on a template T1 image.

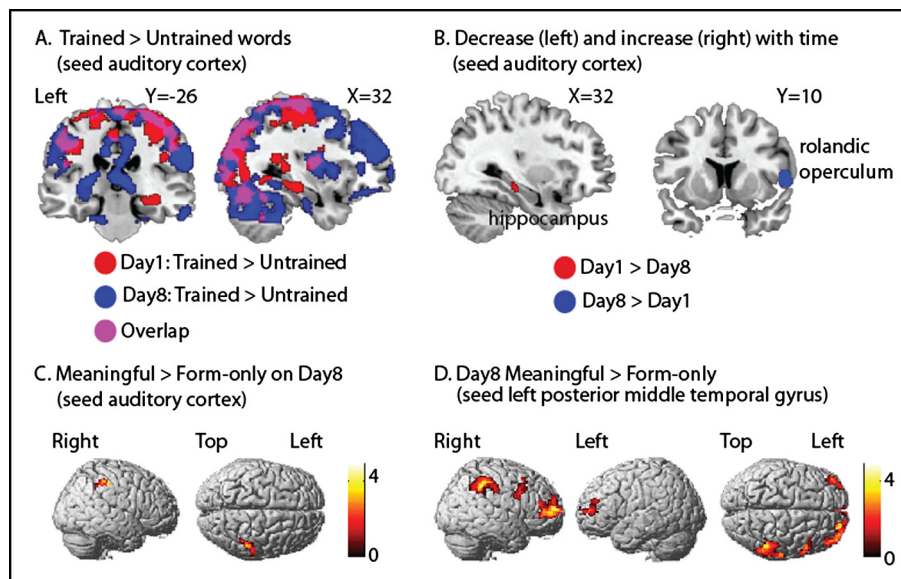


Fig. 3. Functional connectivity. (A) Functional connectivity with seed region in the bilateral auditory cortex. Distributed areas were functionally connected with the auditory cortex on both Day1 and Day8 when the word was correctly recognised relative to when the untrained word was responded to as new. (B) There was, however, a trend for the right hippocampus to connect less to the auditory cortex on Day8 (panel B left), whereas the connection to the right Rolandic operculum increased with time (panel B right). (C) Meaningful novel words showed stronger connectivity with the right inferior parietal lobule relative to Form-only words on Day8. (D) Functional connectivity with seed region in the posterior middle temporal gyrus (pMTG): Meaningful novel words showed stronger connectivity to the multiple neocortical areas relative to Form-only words on Day8. All panels show significant clusters after multiple comparison correction on the cluster level ($p_{FWE} < 0.05$) except for the hippocampal cluster (panel B left) which is depicted at voxel level $P < 0.001$ uncorrected. Colour bars represent the voxel-level t -scores.

that trained novel words will show functional connectivity with memory representation areas, such as the hippocampus for episodic memory, and areas in the distributed cortical areas including the left pMTG in case of the lexical representations. To test this, we ran a PPI analysis comparing functional connectivity that was stronger for the trained words (Form-only + Meaningful) relative to the untrained novel words. Contrast images of the PPI effect for the two sessions were generated for each participant and these contrast images were subjected to second-level analy-

ses. Two-samples t -tests comparing the two groups yielded no significant clusters above the cluster-extent threshold (all clusters $p_{FWE} \geq 0.364$), for both sessions, and the data of the two groups were therefore pooled together for further analyses. For both Day1 and Day8, multiple cortical areas showed a stronger connection with the auditory cortex when recognising trained relative to untrained novel words (Fig. 3A). When we compared the difference in connectivity between Day1 and Day8 (paired t -test), there was a trend towards stronger connectivity between the auditory

cortex and the right hippocampus on Day1 relative to Day8 (peak voxel [32 -26 -12], $p_{\text{svc-hippocampus}} = 0.058$, Fig. 3B left). The reverse contrast (Day8 > Day1) showed stronger connectivity to the right Rolandic operculum (peak voxel [56 10 6], $p = 0.005$, Fig. 3B right).

Next, we were specifically interested in whether the connectivity differed if the novel word was given a meaning (picture or definition), or not. To test this, we calculated the PPI effect between the Meaningful and Form-only conditions. As above, we first generated a contrast image for the PPI effect of the Form-only versus Meaningful conditions for each session per participant, and these images were tested on the second level. Two-samples t -tests comparing the groups yielded no significant clusters (all $p_{\text{FWE}} \geq 0.338$), except for a cluster in the left insula on Day1 that showed a trend to be greater for the Definition group ($p_{\text{FWE}} = 0.095$), thus the data of the two groups were pooled. A comparison between Form-only and Meaningful words did not show any significant differences on Day1, but revealed stronger connectivity on Day8 to the right inferior parietal lobule (IPL; peak voxel [50 -36 48], $p = 0.012$) for Meaningful words compared to Form-only words (see Fig. 3C).

3.2.2.2. Seed from the pMTG

Based on our previous study (Takashima et al., 2014), we assumed that the left pMTG is a lexical hub linking sound to meaning representations that begins to operate for novel words when those words are integrated into the semantic-lexical memory network. We were interested to see if we would find evidence for this in the current data. Since the Meaningful novel words were less abstract than in Takashima et al. (2014), we expected better integration for words in the Meaningful condition relative to the Form-only condition. This difference was hypothesised to be more apparent after a consolidation period on Day8, a delay that was longer than the 24 h in the study by Takashima et al. If the pMTG is acting as a hub region for distributed neocortical representations, we would expect more connectivity with other brain areas for the Meaningful condition relative to the Form-only condition. If consolidation time is required for integration, then we would expect this difference to show up mainly on Day8. To test these hypotheses, a contrast image of the PPI effect with the pMTG as the seed region comparing the Meaningful and Form-only conditions for each day per participant were generated, and these images were tested on the second level. Two-samples t -tests comparing the two groups revealed no Group differences (all $p_{\text{FWE}} \geq 0.862$), and thus the data were pooled across the two groups. We observed no significantly different connectivity strength between Form-only and Meaningful words on Day1. However, and in line with our prediction, the results revealed a stronger distributed connectivity pattern in the neocortex for Meaningful words relative to Form-only words on Day8: the left pMTG connected more to the right IPL/angular/supramarginal gyri and bilateral prefrontal cortex including the left pars triangularis (see Fig. 3D).

In sum, functional connectivity from the auditory cortex revealed that multiple cortical regions were connected when probed with a trained word relative to an untrained word (Table 5, Fig. 3A). With consolidation, there was a trend towards a decrease in connectivity with the right hippocampus, whereas connectivity with the right Rolandic operculum increased (Fig. 3B). The difference between the Form-only and Meaningful conditions became apparent on Day8, with more connectivity between the auditory cortex and the right inferior parietal lobe for the Meaningful condition (Fig. 3C). Furthermore, stronger connectivity with the pMTG for the Meaningful words relative to the Form-only words was observed on Day8 for the right inferior parietal areas and the bilateral prefrontal areas, including the pars triangularis of the left inferior frontal gyrus (Fig. 3D).

4. Discussion

The first aim of the present study was to investigate the memory representations of novel words learned with meanings versus without meanings, and their changes as a function of consolidation. We hypothesised that the hippocampal episodic memory network is more involved during recent retrieval, whereas the neocortical network starts to contribute more strongly after consolidation (Davis & Gaskell, 2009). Furthermore, we hypothesised that this cortical shift would be more pronounced for meaningful novel words. These predictions were partially confirmed. Hippocampal involvement was present on Day1 but was still observable after a delay of one week, albeit less prominently. Neocortical activation was already present on Day1, but a further activation increase in the left IFG (pars triangularis) for the meaningful words was observed after a week of consolidation. Lexical integration was observed on Day8 as suggested by an interference effect during the pause detection task, with a numerically greater effect for the base words of the Meaningful words compared to those of Form-only words. The imaging data showed stronger neocortical activation as well as inter-cortical functional connectivity for Meaningful words compared to Form-only words on Day8, suggesting that integration within the neocortical memory system may be more prominent for Meaningful words after a week of consolidation. On the other hand, Meaningful words also showed heightened hippocampal activation relative to Form-only words on Day8, suggesting that episodic memory was still intact for the Meaningful words after a week's delay. The better memory performance for Meaningful words compared to Form-only words at the delayed test may arise because words with meanings not only have stronger semantic representations, but also maintain intact episodic memory traces.

The second aim of the study was to investigate the stability of semantic memory representations. In the association test, responses in the Form-only and Meaning-same conditions slowed down over time, whereas the RTs for the Meaning-similar condition did not. This is likely because episodic memory representations decay, but semanticised memories are more stable over time.

The third aim of the study was to investigate whether there are differences (behaviourally and neurally) between novel words learned with a pictorial referent and those learned with a definition. Recognition of word-forms was better in the Definition group, although this was mainly driven by better performance on the Form-only words in the Definition group. Free recall of the trained words did not differ between the groups. Imaging results were similar overall for the two groups. The difference between Meaningful and Form-only conditions was less distinct for the Definition group on Day8. We did not find brain activation patterns that were specific to concepts associated with pictures or definitions.

4.1. Memory representations of newly acquired words

Participants utilised multiple available resources and processes in recognising trained words. That is, multiple memory systems appear to work in parallel during recognition of newly learned words. But the degree to which one system is favoured over the other seems to be affected by the time of testing and by the nature of the words.

We may forget actual learning experiences over time, but we do often remember new words and their meanings. Memory consolidation theories suggest a shift from hippocampal to neocortical networks for stabilisation of acquired knowledge (Alvarez & Squire, 1994; McClelland, McNaughton, & O'Reilly, 1995; Squire & Zola-Morgan, 1991; Winocur & Moscovitch, 2011), with sleep as a boosting factor (Diekelmann, 2014). Memory research has

Table 5
Functional connectivity.

	Cluster			Voxel	MNI coordinates (peak)		
	pFWE	Size	z-score		x	y	z
Seed region: auditory cortex							
<i>Day1 Trained vs Untrained</i>							
	<0.001	37,350	5.77	42	-78	12	R middle occipital
			5.61	16	-80	48	R cuneus
			5.51	22	-80	44	R superior occipital
			5.45	36	-18	54	R precentral
			5.44	48	-68	10	R middle temporal
			5.28	26	-60	60	R superior parietal
			5.71	-6	-72	-2	L lingual
			5.28	-4	-52	-4	L cerebellum
	<0.001	370	3.99	14	60	-4	R orbital medial frontal
			3.35	8	18	4	R caudate
	0.042	154	4.22	34	-6	14	R insula
	<0.001	656	4.19	-18	66	6	L superior frontal
			4.1	-24	62	-2	L orbital superior frontal
			4.01	-6	66	6	L medial superior frontal
			3.95	-10	60	-8	L orbital medial frontal
			3.75	-32	62	8	L middle frontal
	0.015	196	4.11	-42	28	30	L inferior frontal pars triangularis
			3.89	-38	30	32	L middle frontal
			3.39	-38	30	16	L inferior frontal
<i>Day8 Trained vs Untrained</i>							
	<0.001	68,783	5.84	-10	-50	-2	L lingual
			5.54	-42	-60	-22	L fusiform
			5.3	-54	-2	0	L superior temporal
			5.76	50	-64	-14	R inferior occipital
			5.73	32	46	18	R middle frontal
			5.72	54	-62	-14	R inferior temporal
			5.57	10	-40	70	R postcentral
			5.48	4	-76	-38	R cerebellum
			5.47	14	-90	28	R cuneus
			5.28	20	-52	56	R superior parietal
<i>Day1 > Day8 Trained vs Untrained</i>							
	0.058 ^a	22	3.76	32	-26	-12	R hippocampus
<i>Day1 < Day8 Trained vs Untrained</i>							
	0.005	288	5.05	56	10	6	R Rolandic operculum
			3.43	54	4	-2	R superior temporal
<i>Day1 Form-only > Meaningful</i>							
	No significant clusters						
<i>Day1 Form-only < Meaningful</i>							
	No significant clusters						
<i>Day8 Form-only > Meaningful</i>							
	No significant clusters						
<i>Day8 Form-only < Meaningful</i>							
	0.012	250	3.75	50	-36	48	R inferior parietal
			3.46	64	-30	44	R supramarginal
Seed region: posterior middle temporal gyrus (pMTG)							
<i>Day1 Form-only > Meaningful</i>							
	No significant clusters						
<i>Day1 Form-only < Meaningful</i>							
	No significant clusters						
<i>Day8 Form-only > Meaningful</i>							
	No significant clusters						
<i>Day8 Form-only < Meaningful</i>							
	<0.001	850	4.41	64	-24	44	R supramarginal
			4.33	50	-32	50	R inferior parietal
	<0.001	945	4.37	40	54	6	R middle frontal
			3.92	46	40	2	R inferior frontal pars triangularis
			3.85	24	52	20	R superior frontal
	0.012	225	4.3	54	14	40	R inferior frontal pars Opercularis
			3.89	52	18	42	R middle frontal
			3.76	60	6	30	R precentral
	0.002	320	4.2	-12	48	6	L medial superior frontal
			4.09	-14	46	10	L anterior cingulum
			3.75	0	66	10	L medial superior frontal

Table 5 (continued)

Cluster		Voxel	MNI coordinates (peak)			
<i>p</i> FWE	Size		z-score	x	y	z
0.006	255	3.98	4	68	8	R medial superior frontal
		3.36	10	54	0	R orbital medial frontal
		3.9	−38	44	24	L middle frontal
		3.54	−44	46	14	L inferior frontal pars triangularis

^a Small volume correction using the hippocampal mask.

revealed a shift from a hippocampus-centred to a neocortically distributed neural representation with consolidation (Frankland & Bontempi, 2005; Takashima et al., 2009). Replicating previous findings (Takashima et al., 2014), the current results with novel word learning followed this pattern. We observed increased neocortical activation, but only for words learned with meanings. Our behavioural finding of a lexical competition effect observed during the pause detection task on Day8, which is believed to be a signature of lexical integration (Gaskell & Dumay, 2003), further supports the notion that the novel words were integrated into the semantic network after a consolidation period. Since the neural patterns observed in the previous study (Takashima et al., 2014) were similar to that tested with a longer delay, the current findings thus extend our understanding of brain areas involved in integration of novel words in that the changes observed after overnight consolidation appear similar to those found after a week's delay. However, if the associated concepts of the novel words are made concrete, lexical integration seems to occur to a numerically stronger degree than when the novel words do not accompany concepts.

The episodic memory system might nevertheless still be efficient in recognising the stimulus if the episodic representation remains vivid. Rich information at encoding helps retain the vividness of the learning experience. As the hippocampus is a well-known structure involved in encoding and retrieving episodic memories (Nadel & Moscovitch, 1997; Squire, 1992), also for novel words (Breitenstein et al., 2005; Davis et al., 2009), we can infer that activation in the hippocampus is a signature of episodic memory retrieval. We observed the bilateral hippocampus to be active during recognition of the trained novel words on Day1. Activation was still observable on Day8 but less so, especially for the Form-only words. There was a trend for the functional connectivity between the hippocampus and the auditory cortex to decrease with time, further supporting the idea that the information processing route through the hippocampal episodic memory system decreases with consolidation. Furthermore, the activation of the left precuneus decreased with time for both Meaningful and Form-only words. The precuneus is known to be a part of the episodic/autobiographical memory network (Cavanna & Trimble, 2006), and is more active at retrieval than encoding (Daselaar et al., 2009; Huijbers et al., 2012). Thus the activation of this area is likely to reflect episodic memory retrieval, consistent with the idea that the episodic nature of the recognised novel words declined with time.

It was not the case, however, that only episodic memory retrieval took place during the recent memory test, and only semantic memory representations were activated at the delayed memory test. More activation in the hippocampus for Meaningful words relative to Form-only words on Day8 suggests that rich information at encoding helps learners retain the vividness of the learning experience, and that episodic memory is still involved in recognition of the trained words on Day8.

On Day8, multiple cortical areas became more active for Meaningful compared to Form-only words. The areas included the hypothesised areas in the posterior temporal/parietal areas, including the left angular gyrus, posterior middle temporal gyrus,

supramarginal gyrus, and the inferior parietal lobe. These areas are reported to be active during learning of phonologically presented novel words (Breitenstein et al., 2005; McNealy, Mazziotta, & Dapretto, 2006; Paulesu et al., 2009), visually presented words (Mestres-Missé et al., 2008), and also when previously trained novel words are remembered (Cornelissen et al., 2004; Hultén, Laaksonen, Vihla, Laine, & Salmelin, 2010). A recent study showed increased activation of the left angular gyrus when the associated information was semantically rich for both novel and familiar words, suggesting a role for this area in semantic retrieval (Ferreira, Göbel, Hymers, & Ellis, 2015). For processing of familiar words, the left angular gyrus is often reported to be involved (reviewed in Binder et al., 2009; Price, 2010).

The posterior middle temporal gyrus has been proposed to play a role in integrating auditory, visual, and semantic information (Gow, 2012). In the previous study, we found that participants who showed greater integration of novel words also recruited the left posterior middle temporal gyrus to a greater degree (Takashima et al., 2014). The current results showed more engagement of this area for Meaningful words than Form-only words on Day8, in line with the behavioural integration effect (competition effect found during the pause detection task) that was numerically greater for Meaningful than Form-only condition. Another candidate area in the temporal cortex that serves as a semantic hub is the anterior temporal lobe (Patterson et al., 2007). Our results showed more activation in the right anterior temporal lobe for the Meaningful compared to Form-only condition overall, suggesting that upon hearing a word that is meaningful, the concept is probably retrieved as well.

The left inferior frontal gyrus, specifically the pars triangularis, showed a differential activation pattern. This area showed elevated activation for Meaningful novel words over Form-only words in general, especially after a delay on Day8. This was also the area that showed greater functional connectivity with the auditory cortex for Trained compared to Untrained novel words, and greater functional connectivity with the left posterior middle temporal gyrus for Meaningful compared to Form-only words on Day8. In a study where participants were exposed to a stream of speech sounds, the left inferior frontal gyrus was more active for repeated presentation of trisyllabic words embedded in the stream compared to a combination of syllables that was not repeated (McNealy et al., 2006). The authors of this study proposed that this area is involved in word segmentation. Mestres-Missé et al. found that when participants had to deduce the meaning of the novel word in a sentence context, this area in the inferior frontal gyrus was engaged (2008), which led the authors to propose that this area controls the retrieval of semantic knowledge. Our results are in line with the semantic retrieval account as we found the pars triangularis to be more involved in recognition of Meaningful novel words relative to Form-only words.

To summarise, our findings indicate that the memory representation of a novel word does not shift completely from the episodic memory system to the semantic memory system through consolidation, at least not within the time span that we tested (one week). Rather, the multiple memory systems continue to work in parallel,

but contribute to retrieval to a different degree depending on the richness of the lexical information.

4.2. Stability of semantic representations

The RT results from the association test provide insight into the stability of semantic memory representations. Participants had to choose the associated information upon hearing the trained novel word. Retrieval of the associated referent (picture/definition) when cued with the phonological word-form can be executed via the episodic system if audio-visual material used in the test is identical to that during training, whereas if the visual presentation of the referent is altered, abstracted conceptual representations need to be activated to perform the task. The RT data from the association test suggest that recruitment of the semantic memory system may be slower in general, as overall longer RTs were observed for the Meaning-similar condition compared to the Meaning-same condition. Comparison across days, however, showed a slowing down for the Form-only and Meaning-same conditions but not for the Meaning-similar condition. This suggests that the episodic memory trace weakened over time, leading to slower RTs on Meaning-same trials on Day8 relative to Day1 whereas the semantic representations that were accessed on Meaning-similar trials remained stable over time. Thus, these data corroborate the idea that novel episodic memory traces are more transient in nature than novel semantic memories.

4.3. Learning novel words with and without meanings

On the behavioural level, both recognition and free recall performance was better for the Meaningful words compared to the Form-only words.

The imaging data revealed that whereas there was no difference between Meaningful and Form-only words on Day1, neocortical activation was greater for the Meaningful words (compared with Form-only words) on Day8. Functional connectivity analyses also showed a distributed neocortical connectivity pattern that was stronger for Meaningful words compared to Form-only words on Day8. At the same time, persistent hippocampal involvement was also found for Meaningful words on Day8. These results suggest that the semantic richness of the memory representations for the Meaningful words enabled the use of multiple memory systems, both episodic and semantic, even after a week's delay, leading to better memory performance for these words compared to the Form-only words.

4.4. Integration of novel words

Stabilisation of discrete novel word representations is one thing, but integration of the novel information into the existing lexicon is another. A novel word representation will only start to interact with neighbouring words in an automated manner if the representation becomes integrated into the cortical network (Gaskell & Dumay, 2003; Leach & Samuel, 2007). Once integrated, closely related elements, such as words with similar phonology, similar orthography, and related meanings become linked with the novel word and start affecting the recognition process of associated words. Prior behavioural studies have exploited this phenomenon to measure integration of novel words into the mental lexicon (Bakker et al., 2014; Davis & Gaskell, 2009; Gaskell & Dumay, 2003). To show an interference effect during the pause detection task, a novel word needs to establish a connection with its existing phonological neighbours in order to affect their recognition speed. The slowing down of responses to base words of trained novel words in the current data suggests that this integration process had indeed taken place.

Unlike our previous finding that Form-only words showed greater interference effects than Meaningful words (Takashima et al., 2014), in the current study, using the same pause detection task, Meaningful words showed the stronger interference effect. There are several possible reasons for this difference. Time is a factor, because in the current study we tested a week later rather than 24 h later, as in Takashima et al. (2014). Previous studies suggest that both time and repeated exposure lead to greater interference effects (Bakker et al., 2014; Lindsay & Gaskell, 2013). However, this does not explain the reversal of the difference between Form-only words and Meaningful words. The more likely explanation, in our opinion, relates to how the novel words were encoded. In Takashima et al. (2014) as well as many of the previous studies showing word-form integration, participants' attention during training was focused on how the novel word sounded (Bakker et al., 2014; Davis et al., 2009; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003; Henderson, Weighall, & Gaskell, 2013; Lindsay & Gaskell, 2013; Takashima et al., 2014; Tamminen & Gaskell, 2008). Furthermore, in Takashima et al. (2014), we contrasted the consolidation of words with and without referents, and the visual referents were abstract shapes which made the audio-visual association more arbitrary in nature. In the current study, training drew attention to the association between the word-form and its referent, and all referents were existing objects. As this made the stimuli conceptually easier to assimilate with existing semantic knowledge, and because the learning of the relationship between the word and the meaning was encouraged in the current study, these factors may have advanced the lexicalisation process for Meaningful words over Form-only words.

4.5. Difference in remembering novel words with picture referents and verbal descriptions

We did not observe a systematic behavioural benefit of learning word meanings for those words learned with pictures or those learned with definitions. Whereas recognition performance of the words themselves was better for the Definition group, the number of words recalled did not differ between groups for the meaningful condition and at the free recall test. Regarding the imaging results, our input modality was auditory, and activation was observed in the bilateral auditory cortex for all conditions in both groups. When the activation patterns for Meaningful words were compared between the Picture and Definition groups, we did not find any significant clusters indicating greater activation for one group than the other. This suggests that if the input modality of the word is kept constant but stimulus manipulation is on the level of associated information (word meaning), then brain activation patterns seem not to differ from each other.

5. Conclusion

Novel words richer in information at encoding were better retained at a delay of one week. We showed that both episodic and semantic memory systems were utilised during recognition of these novel words, but the involvement of the two systems changed as a function of time and the amount of associated information. RT changes in the association test indicated decay of the episodic memory whereas the semantic memory trace seemed to be strengthened with time, as suggested by the pattern of brain activity in the distributed neocortical areas during recognition of meaningful novel words. These two memory systems were nonetheless engaged during retrieval at both delays, especially for the words learnt with meanings. We suggest that meaningful words are remembered better because their retrieval benefits more from these complementary memory systems.

Acknowledgments

This research is funded by The Netherlands Organization for Scientific Research (NWO) Brain and Cognition grant no. 433-09-239. We thank Paul Gaalman for technical support.

References

- Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences of the United States of America*, 91(15), 7041–7045.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. L. D. Consortium.
- Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, 73(1), 116–130.
- Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, 15(11), 527–536. <http://dx.doi.org/10.1016/j.tics.2011.10.001>.
- 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. *Cerebral Cortex*, 19(12), 2767–2796. <http://dx.doi.org/10.1093/cercor/bhp055>.
- Boersma, P., & Weenink, D. (2012). Praat: Doing phonetics by computer [computer program]. Version 5.3.23 from <http://www.praat.org/>.
- 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.
- Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, 129(3), 564–583. <http://dx.doi.org/10.1093/brain/awl004>.
- Clay, F., Bowers, J. S., Davis, C. J., & Hanley, D. A. (2007). Teaching adults new words: The role of practice and consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), 970–976. <http://dx.doi.org/10.1037/0278-7393.33.5.970>.
- Cornelissen, K., Laine, M., Renvall, K., Saarinen, T., Martin, N., & Salmelin, R. (2004). Learning new names for new objects: Cortical effects as measured by magnetoencephalography. *Brain and Language*, 89(3), 617–622. <http://dx.doi.org/10.1016/j.bandl.2003.12.007>.
- Damasio, A. R. (1989). Time-locked multiple retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33(1–2), 25–62. [http://dx.doi.org/10.1016/0010-0277\(89\)90005-X](http://dx.doi.org/10.1016/0010-0277(89)90005-X).
- Daselaar, S. M., Prince, S. E., Dennis, N. A., Hayes, S. M., Kim, H., & Cabeza, R. (2009). Posterior midline and ventral parietal activity is associated with retrieval success and encoding failure. *Frontiers in Human Neuroscience*, 3. <http://dx.doi.org/10.3389/fnhum.009.013.2009> (Original Research).
- Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, 21(4), 803–820. <http://dx.doi.org/10.1162/jocn.2009.21059>.
- Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1536), 3773–3800. <http://dx.doi.org/10.1098/rstb.2009.0111>.
- Diekmann, S. (2014). Sleep for cognitive enhancement. *Frontiers in Systems Neuroscience*, 8, 46. <http://dx.doi.org/10.3389/fnsys.2014.00046>.
- Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, 18(1), 35–39. <http://dx.doi.org/10.1111/j.1467-9280.2007.01845.x>.
- Ferreira, R. A., Göbel, S. M., Hymers, M., & Ellis, A. W. (2015). The neural correlates of semantic richness: Evidence from an fMRI study of word learning. *Brain and Language*, 143, 69–80. <http://dx.doi.org/10.1016/j.bandl.2015.02.005>.
- Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, 6(2), 119–130.
- Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, 89(2), 105–132. [http://dx.doi.org/10.1016/S0010-0277\(03\)00070-2](http://dx.doi.org/10.1016/S0010-0277(03)00070-2).
- Gow, D. W. (2012). The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. *Brain and Language*, 121(3), 273–288. <http://dx.doi.org/10.1016/j.bandl.2012.03.005>.
- Graves, W. W., Binder, J. R., Desai, R. H., Conant, L. L., & Seidenberg, M. S. (2010). Neural correlates of implicit and explicit combinatorial semantic processing. *Neuroimage*, 143, 638–646. <http://dx.doi.org/10.1016/j.neuroimage.2010.06.055>.
- Hayasaka, S., & Nichols, T. E. (2003). Validating cluster size inference: Random field and permutation methods. *Neuroimage*, 20(4), 2343–2356.
- Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2013). Online lexical competition during spoken word recognition and word learning in children and adults. *Child Development*, 84(5), 1668–1685. <http://dx.doi.org/10.1111/cdev.12067>.
- Henderson, L. M., Weighall, A. R., & Gaskell, M. G. (2013). Learning new vocabulary during childhood: Effects of semantic training on lexical consolidation and integration. *Journal of Experimental Child Psychology*, 116(3), 572–592. <http://dx.doi.org/10.1016/j.jecp.2013.07.004>.
- 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. <http://dx.doi.org/10.1016/j.cognition.2003.10.011>.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.
- Huijbers, W., Vannini, P., Sperling, R. A., Pennartz, C. M., Cabeza, R., & Daselaar, S. M. (2012). Explaining the encoding/retrieval flip: Memory-related deactivations and activations in the posteromedial cortex. *Neuropsychologia*, 50(14), 3764–3774. <http://dx.doi.org/10.1016/j.neuropsychologia.2012.08.021>.
- Hultén, A., Laaksonen, H., Vihla, M., Laine, M., & Salmelin, R. (2010). Modulation of brain activity after learning predicts long-term memory for words. *The Journal of Neuroscience*, 30(45), 15160–15164. <http://dx.doi.org/10.1523/JNEUROSCI.1278-10.2010>.
- Kapnola, E. C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, 134, 85–99. <http://dx.doi.org/10.1016/j.cognition.2014.09.007>.
- Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, 55(4), 306–353. <http://dx.doi.org/10.1016/j.cogpsych.2007.01.001>.
- Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(2), 608–622. <http://dx.doi.org/10.1037/a0029243>.
- 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. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, 58(1), 25–45. <http://dx.doi.org/10.1146/annurev.psych.57.102904.190143>.
- Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology*, 11(2), 194–201.
- 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. *Psychological Review*, 102(3), 419–457.
- McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: Neural mechanisms underlying speech parsing. *The Journal of Neuroscience*, 26(29), 7629–7639. <http://dx.doi.org/10.1523/jneurosci.5501-05.2006>.
- Mestres-Missé, A., Càmarà, E., Rodríguez-Fornells, A., Rotte, M., & Münte, T. F. (2008). Functional neuroanatomy of meaning acquisition from context. *Journal of Cognitive Neuroscience*, 20(12), 2153–2166. <http://dx.doi.org/10.1162/jocn.2008.20150>.
- Meyer, K., & Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neurosciences*, 32(7), 376–382. <http://dx.doi.org/10.1016/j.tins.2009.04.002>.
- Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., & Rosenbaum, R. S. (2006). The cognitive neuroscience of remote episodic, semantic and spatial memory. *Current Opinion in Neurobiology*, 16(2), 179–190.
- Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, 7(2), 217–227.
- 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. *Nature Reviews Neuroscience*, 8(12), 976–987.
- Paulesu, E., Vallar, G., Berlinger, M., Signorini, M., Vitali, P., Burani, C., ... Fazio, F. (2009). Supercalifragilisticexpialidocious: How the brain learns words never heard before. *Neuroimage*, 45(4), 1368–1377. <http://dx.doi.org/10.1016/j.neuroimage.2008.12.043>.
- Penny, W. D., Holmes, A. P., & Friston, K. J. (2003). Random effects analysis. In R. S. J. Frackowiak, K. J. Friston, C. Frith, R. Dolan, C. J. Price, S. Zeki, J. Ashburner, & W. D. Penny (Eds.), *Human brain function* (2nd ed., 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. *Magnetic Resonance in Medicine*, 55(6), 1227–1235. <http://dx.doi.org/10.1002/mrm.20900>.
- Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, 1191, 62–88. <http://dx.doi.org/10.1111/j.1749-6632.2010.05444.x>.
- Simmons, W. K., & Barsalou, L. W. (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology*, 20(3–6), 451–486. <http://dx.doi.org/10.1080/02643290342000032>.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99(2), 195–231.
- Squire, L. R., & Knowlton, B. (1994). Memory, hippocampus, and brain systems. In M. Gazzinga (Ed.), *The cognitive neurosciences. Neuropsychology in geriatric neurology*. Cambridge, MA: MIT Press.
- Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, 253(5026), 1380–1386.
- Szmalc, A., Page, M. P. A., & Duyck, W. (2012). The development of long-term lexical representations through Hebb repetition learning. *Journal of Memory and Language*, 67(3), 342–354. <http://dx.doi.org/10.1016/j.jml.2012.07.001>.
- Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *Neuroimage*, 84, 265–278.
- Takashima, A., Nieuwenhuis, I. L. C., Jensen, O., Talamini, L. M., Rijpkema, M., & Fernández, G. (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. *The Journal of Neuroscience*, 29(32), 10087–10093.

- Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *The Quarterly Journal of Experimental Psychology*, 61(3), 361–371. <http://dx.doi.org/10.1080/17470210701634545>.
- Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *The Quarterly Journal of Experimental Psychology*, 66(5), 1001–1025. <http://dx.doi.org/10.1080/17470218.2012.724694>.
- 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. *The Journal of Neuroscience*, 30(43), 14356–14360. <http://dx.doi.org/10.1523/jneurosci.3028-10.2010>.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... 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. <http://dx.doi.org/10.1006/nimg.2001.0978>.
- van der Ven, F., Takashima, A., Segers, E., & Verhoeven, L. (2015). Learning word meanings: Overnight integration and study modality effects. *PLoS One*, 10(5), e0124926. <http://dx.doi.org/10.1371/journal.pone.0124926>.
- Vandenberghe, R., Wang, Y., Nelissen, N., Vandenbulcke, M., Dhollander, T., Sunaert, S., & Dupont, P. (2013). The associative-semantic network for words and pictures: Effective connectivity and graph analysis. *Brain and Language*, 127(2), 264–272. <http://dx.doi.org/10.1016/j.bandl.2012.09.005>.
- Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society*, 17(5), 766–780. <http://dx.doi.org/10.1017/S1355617711000683>.