Memorization, lexicalization, semanticization, and consolidation of novel words in L2.

Giacomo Tartaro

Supervisors: James McQueen [1,2], Atsuko Takashima [1,2]

1 Donders Centre for Cognitive Neuroimaging, Nijmegen, Netherlands

2 Donders Centre for Cognition, Nijmegen, Netherlands

Keywords: word learning, second language acquisition, memorization, lexicalization, semanticization, consolidation, hippocampus, pMTG, IFG

Abstract

The memorization, lexicalization, semanticization, and consolidation of novel words have been mainly investigated in the first language (L1). This research explored this processes in a second language (L2) and the influence on these processes of individual differences: length of stay in an L2 environment, proficiency and vocabulary size of L2 on this four processes. Italian-English (late) bilinguals learned a set of 40 English pseudo-word/picture associations on the first day of training (remote condition), and after two days they repeated the procedure with a different set (recent condition) and then performed the testing phase. Only the vocabulary size correlated with the memorization process in the training phase. The outcome of a lexical competition task (pause detection) showed a strong competition effect for the remote condition but not for the recent condition. A primed lexical decision task showed a significant priming effect with the two conditions merged. A recognition memory task in the fMRI showed activation for the remote condition of the inferior frontal gyrus, an area which is thought to unify semantic information of different modalities. Hippocampus, involved in episodic memory and at the first stage of encoding of novel words, was active in both conditions. Overall, the consolidation process seems to not have a one-to-one correspondence with lexicalization and semanticization and a higher dependency on the hippocampus for both the Remote and Recent condition. The lexicalization of the word-form is totally in line with previous literature in L1 but not the semanticization of the meaning which nevertheless shows a trend in line with the previous literature in L1. Finally, word-learning seems to not differ between L1 and L2, at the behavioral level, with some beneficial effects of the vocabulary size in L2 on the memorization process. Also, sleep seems to be beneficial for semanticization and lexicalization effects to arise at the behavioral level.

Corresponding author: Giacomo Tartaro, email: tartarogiacomo@gmail.com

1 Introduction

In our daily life, we constantly encounter and learn new words. But what does it mean to learn a word? It means to learn its significance, that is the meaning, and to learn its form, that is its pronunciation and spelling. For example, we could learn that an “appet” is a weird kind of bird. So one hand we have to memorize the word-form “appet”, with all its linguistic properties (sound, spelling, etc.), and one the other hand we have to learn its meaning, “a weird kind of bird”, with all its semantic connection (it is an animal, it flies etc.). But does learning a concept differs from learning a word-form and which are the brain areas involved? To learn new words is particularly challenging when we face new languages and we need to learn several new words. But is word learning a process which differs if it occurs in our native language (L1) or in a foreign language (L2)? Or is it always the same, regardless of which language we are learning the new words? Moreover, do individual differences play a role in word learning? Does the proficiency in L2 help to learn new words in that language? If we know several words in L2 it will be easier for us to learn new words in L2 or not? Is word learning influenced by our usage and immersion in a language, as when we go to live abroad? In order to answer all these questions, in this study, we tracked the acquisition of the form and meaning of new words in a second language using both behavioral and neural measures of learning and collected measurements of proficiency, dictionary size, and length of stay.

1.1 Overall perspective

In order to disentangle the learning process of word-forms and concepts, two different tasks have been used in this experiment: lexical competition - pause detection task and primed lexical decision task (Fig. 1).

The pause detection task can verify if a word had been lexicalized, hence, incorporated in the mental lexicon. When this occurs, lexical competition arises which slows down the reaction times (RTs) in the pause detection task (Gaskell & Dumay, 2003a). In a similar vein, semantic priming verifies the semanticization of a word, hence, the insertion of the meaning of the word in the mental lexicon at the conceptual level (Verhoeven, Ven, Takashima, & Segers, 2015). If a concept has been semanticized, when we hear the word which expresses it, the concept will prime semantically related words in a lexical decision task, with faster RTs.

This two effects arise when the words and their related concepts have been inserted in the mental lexicon, namely the "mental dictionary" which contains information regarding a word's meaning and linguistic properties (phonology, orthography, morphology, syntactic characteristics).

It is important to underline that the memorization of a novel word is not sufficient to show the effects of lexicalisation and semanticization previously described. Sleep is thought to be beneficial for the arising of the interaction between the novel word and the known words at both the meaning (Clay, Bowers, Davis, & Hanley, 2007;

Tamminen & Gaskell, 2013; Verhoeven et al., 2015) and word-form level (Bakker, Takashima, van Hell, Janzen, & McQueen, 2014). The role of sleep can be explained by the complementary learning system (Davis & Gaskell, 2009). When a novel word is learned, its memory representation is thought to undergo a process of consolidation with a shift from episodic and hippocampus-dependent memory to a semantic memory associated with distributed neocortical areas. When this shift occurs, the lexicalization and semanticization effects could be observed at the behavioral level. Sleep has been found to be one of the driving factors

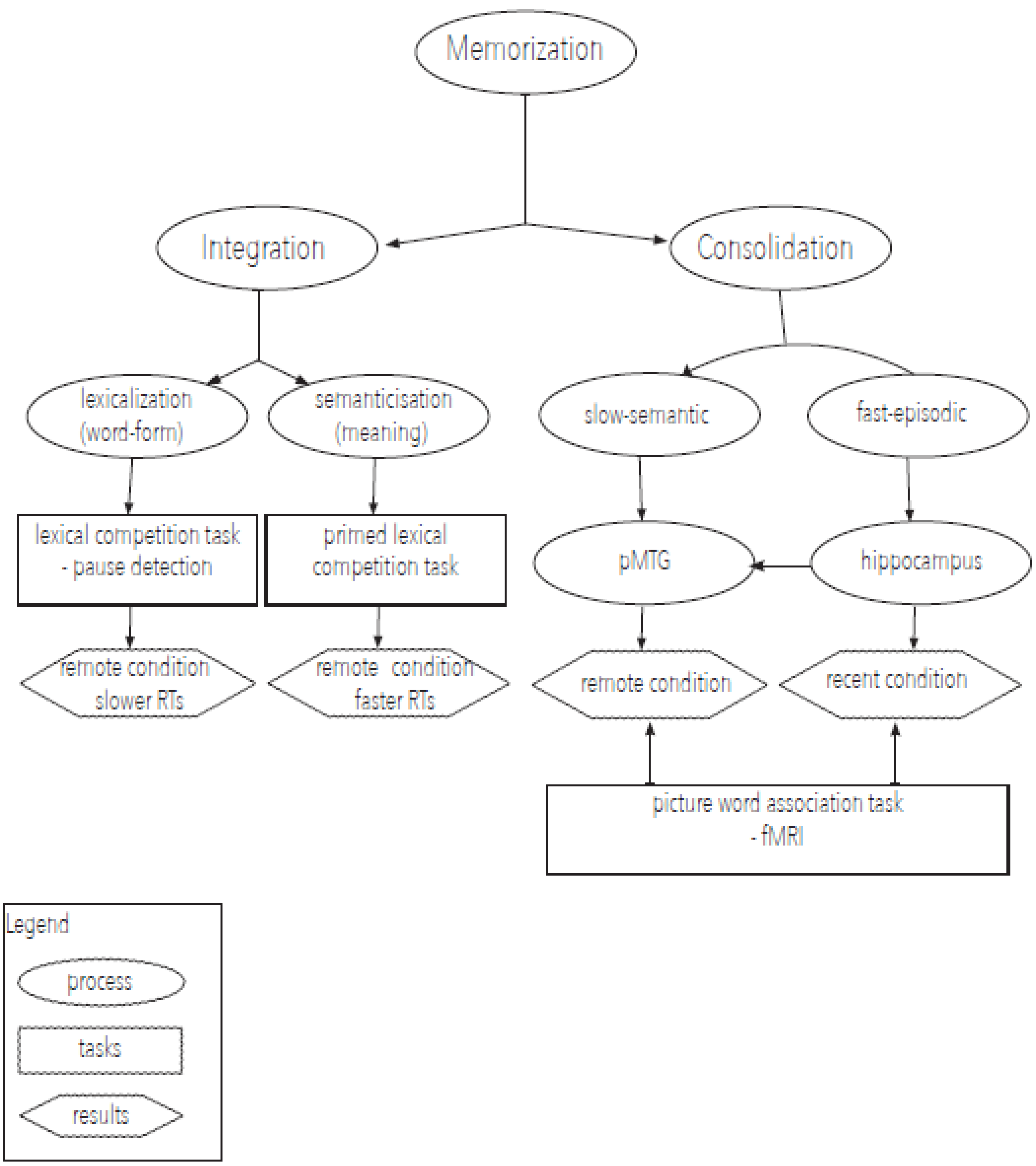

Fig 1 – An overall perspective of the processes, tasks, and areas involved in this research. For consolidation is meant the changes in the brain. For Integration is meant the insertion of a word in the mental lexicon both at the word-form level and meaning level. Lexicalization refers to the integration of the word-form only and for semanticization the integration of the word at the meaning level only.

(Dumay & Gaskell, 2007; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2002) for the integration of the novel word in the brain. The areas involved in the retrieval and usage of consolidated words are thought to be the inferior frontal gyrus (IFG) and the posterior middle temporal gyrus (pMTG) as shown by some researches (Krieger-Redwood, Teige, Davey, & Hymers, 2015; Price, 2012; Ripollés et al., 2017) and also as assumed, in the memory-unification-control model (Hagoort, 2013). The IFG, is thought to unify semantic information of different modalities and pMTG to be responsible for the retrieval of lexical-syntactic information from memory (Krieger-Redwood et al., 2015; Price, 2012; Ripollés et al., 2017; Willems, Özyürek, & Hagoort, 2007, 2008). In addition, the Dual Stream Model (Hickok & Poeppel, 2004) claims that the pMTG is responsible for the lexical interface between the phonological network and the conceptual network, hence, the area responsible for connecting words to their concept.

To investigate the consolidation process we conducted a recognition memory task (picture-word association) in the fMRI scanner, hypothesizing pMTG activation for words which have underwent two nights of sleep and consolidation (Remote condition) and hippocampal activation for words learned the same day of testing (Recent condition). Moreover, we also expected lexicalization effects to arise only for the Remote condition both at the meaning level and word-form level. In addition we investigated the role of length of stay, proficiency and vocabulary size on this processes, which could show some interaction. All four processes (memorization, lexicalization, semanticization and consolidation, see Fig. 1) are well established in the first language (L1) but they have been rarely investigated in a second language (L2). The aim of this research is to examine the similarities and differences between L1 and L2 regarding these processes, hence, a population of Italian-English (late) bilinguals was tested and specific pseudo-words were created to avoid the usage of L1 words focusing selectively on the L2 lexicon.

1.2 Consolidation - the complementary learning systems account

Davis and Gaskell (Gaskell & Dumay, 2003a; Mcclelland, Mcnaughton, & Reilly, 1995; Verhoeven et al., 2015) applied the complementary learning systems account (CLS) of McClelland et al. (1995) to adult word learning, suggesting it occurs via two complementary learning systems.

The first learning system rapidly acquires a novel word and stores it as an episodic memory trace. This memory system is supported by the medial temporal lobe, including the hippocampus and it is independent of the already existing network of word representations. This network is supposed to be used when remembering the experience of learning a novel word (Lindsay & Gaskell, 2010).

The second learning system is responsible for the post-acquisition learning processes with a shift towards a more semanticized and lexicalized coding of the memory representation. This semantic memory system is distributed over a neocortical network and is used when we retrieve the semantic information associated with the word (Binder & Desai, 2011; Martin & Chao, 2001; Patterson, Nestor, & Rogers, 2007).

Also, the duration of the memory in the two systems differs: the episodic hippocampal memory decays, whereas the semantic memory is more stable over time. Once consolidation has occurred, with the shift from hippocampal to cortical networks, it should lead to the formation of new relations between novel and old knowledge at both meaning level and word-form level. Consolidation should be associated with the appearance of semantization and lexicalization effects at the behavioral level.

The areas involved in the cortical network of the consolidated words are thought to be, among the others, IFG and pMTG. The IFG, is involved in the unification process of semantic information of different modalities and pMTG retrieves lexical-syntactic information from memory (Willems et al., 2007, 2008). There are also other areas involved in the usage of consolidated words but these areas are related to other features of the words, such as phonetic features and spelling for example, which are beyond the scope of this research.

Sleep has been found to be an important factor for the consolidation of memory in general (Casey et al., 2016; Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994; Mcclelland et al., 1995; R Stickgold, 1998; Robert Stickgold, 2005; Walker, Brakefield, Morgan, Hobson, & Stickgold, 2002; Walker & Stickgold, 2004; Walker M.P., 2003) and as well for the integration of the words in the mental lexicon, which has been found in several studies only after a delay of 24 h or a week both at the word-form level (Bakker et al., 2014; Gaskell & Dumay, 2003b). and conceptual level (Bakker et al., 2014; van der Ven, Takashima, Segers, & Verhoeven, 2017; Verhoeven et al., 2015) Sleep has been found to be one of the driving factors for integration (Dumay & Gaskell, 2007; Tamminen et al., 2002) although some studies suggest that sleep may not be important and that integration can occur directly after training both at the word form level (Kapnoula, Packard, Gupta, & McMurray, 2015; Szmalec, Page, & Duyck, 2012; Lindsay, S., & Gaskell, M. G. 2013) and meaning level (Bordag et al., 2015; Kapnoula & McMurray, 2016; Kapnoula et al., 2015; Lindsay & Gaskell, 2013.; Oppenheim, 2018; Szmalec et al., 2012).

We have therefore two sides of the same coin: the consolidation at the brain level on one side and the integration of the words in the mental lexicon on the other side. The latter one can be furthermore halved in lexicalization at the word-form level and semanticization at the conceptual level (see Fig. 1).

In L1 both consolidation and integration effects have been found already after one night of consolidation (Davis, Di Betta, Macdonald, & Gaskell, 2009), therefore, in this research we asked if consolidation and integration follow the same time-pattern in L2. In order to investigate this question, participants had to learn a set 40 picture-word associations the first day of the experiment (Remote condition) and then, two days after, repeated the same procedure with a different set (Recent condition) and the testing phase (see Fig. 2). The test phase consisted in three different tasks: lexical competition task – pause detection, primed lexical decision task, and one task in the fMRI scanner. The task in the scanner was to correctly associate a picture to its word-form among four options. We chose this task because we are interested in the brain areas responsible for the connection of

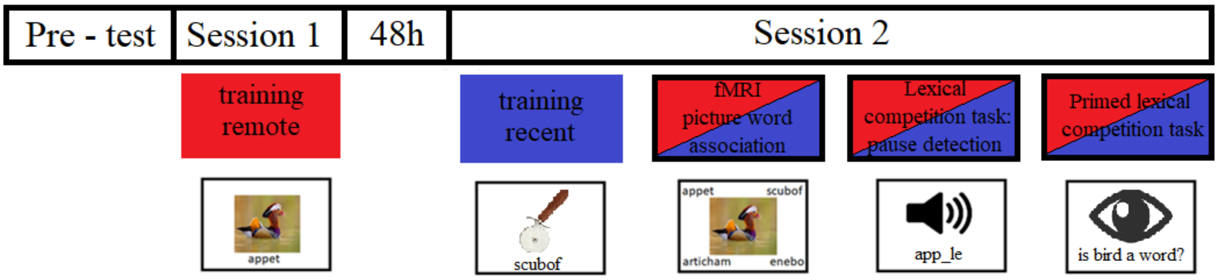

Fig2 – An overall perspective of the whole procedure.

the word-form to its meaning. We chose two night of consolidation in order to enhance as much as possible the consolidation process without a substantial memory loss for the words of the remote condition (as could occur be after a whole week of delay between the two phases). We did that in order to have enough imaging data also for the remote condition. Hence, the words of the remote condition underwent two nights of consolidation and should show the integration behavioral effects and as well the consolidation effect in the brain.

We expected to find hippocampal activation for words of the recent condition and pMTG activation for words of the remote condition in line with the complementary learning systems account (McClelland, McNaughton, & O'Reilly, 1995) and also with the system consolidation theory (Squire, Stark, & Clark, 2004). We hypothesized pMTG activation since is thought to be a hub region for connecting the audio and visual representations of the consolidated words (Gow, 2012; Hagoort, 2013; Hickok & Poeppel, 2004; Krieger-Redwood et al., 2015; Price, 2012; Ripollés et al., 2017; Takashima, Bakker, van Hell, Janzen, & McQueen, 2014, 2017).

Since the pMTG, together with the IFG, is responsible for the connection of the word-form to the conceptual meaning we also investigated at the behavioral level if the integration of concept and word-form occurred. In that way, we can ensure that pMTG and IFG are responsible for the creation of interaction between the known words and the learned words both at the conceptual and word-form level and that consolidation is linked to integration also in L2.

In order to investigate the integration at the word-meaning level, we used a primed lexical decision task and to examine the integration at the word-form level we used a lexical competition task (pause detection).

1.3 Lexicalization - Lexical competition task (pause detection)

One of the strongest pieces of evidence for the occurrence of lexicalization of a novel word is that the new word plays a role in lexical competition (Gaskell & Dumay, 2003b), which is the inhibitory interaction among words in speech comprehension. If people learn a new word, for example, "*appet*", the presence of this new word in the mental lexicon will delay the recognition of similar existing words such as "*apple*". This is evidence of lexicalization because the assimilation of a word in the mental lexicon has effects on the processing of already existing words in the lexicon and it is widely established (Luce & Pisoni, 1998; W. D. Marslen-Wilson, 1987).

Several pieces of evidence for lexical competition has been found, for onset matching words (W. Marslen-Wilson & Zwitserlood, 1989.) and more generally for overlapping words such as “mess” and “domestic” (Cluff & Luce, 1990; Dumay, Frauenfelder, & Content, 2000; Goldinger, Luce, & Pisoni, 1989; McQueen, Norris, & Cutler, 1994.; Norris, McQueen, & Cutler, 1995; Vroomen & de Gelder, 1995).

The reason why this competition arises is the activation of multiple lexical candidates given the same portion of the input. For example, the string “sta” will active “stage” and “state”, which will compete with each other. It has also been shown that onset-matching words (as in the example) are stronger competitor compared to rhyme competitor (as “lizard” vs. “wizard”), whose competition also arise later (Dahan, Magnuson, & Tanenhaus, 2001). Hence, in order to enhance lexical competition, we created pseudo-words which diverge from the existing base word in the last syllable, resembling the onset-matching words paradigm.

We wanted to enlarge as much as possible the size of the competition effect because it is renowned to be small, usually the RTs slow down between 20 and 40 ms on average (Gaskell & Dumay, 2003a; McQueen et al., 1994.). Hence, it can be difficult to detect such a small variation and we wanted to maximize as much as possible our opportunity to find it.

The competition of onset-matching words occurs until the divergence point (i.e. the point where the two or more words start to differ from each other) and it has been demonstrated that pause detection is a good online tool to investigate lexical activity, showing that the speed of the pause detection is dependent on the overall amount of lexical activity caused by the preceding speech (Mattys & Clark, 2002). Hence, the moment with the higher competition occurs at the divergence point. For this reason, as in Gaskell’s research (Gaskell & Dumay, 2003b), we choose this task and inserted the artificial pause at the divergent point.

Lexical competition has also been found to be stronger for words with smaller phonological neighborhood (Frauenfelder & Peeters, 1990). The reason why this occurs can be easily explained. For example, we can compare “tale” and “apple”. The word “apple” does not have any phonological neighborhoods, on the other hand, “tale” has several of them (“take”, ”tape”, ”tame”, “table”,” tailor”, “taint”, “tail”). If we learn a new similar word to “tale”, such as “tace”, we change the number of competitors from seven to eight, which will have a small impact in the RTs of the response. On the contrary, if we learn a word similar to “apple”, as for example “appet”, we increase the number of competitors from zero to one and this is going to have a bigger impact on the RTs of the response. For this reason, we choose words with zero or few phonological neighborhoods in order to enhance lexical competition as much as possible after learning of the novel word.

Since other researchers have already shown that lexical competition can arise between different languages (Costa, Miozzo, & Caramazza, 1999; Weber & Cutler, 2004) we chose English words which have no cognates or false friends in Italian. In this way we were able to avoid competition among Italian neighboring words and to focus on the English lexicon only. We focused on the English lexicon only, because we wanted to observe if there are differences in word learning between L1 and L2 taking out any possible L1 influence on L2.

It is important to underline that the lexical competition which is measured with the pause detection task is regardless of the knowledge of the meaning of the word, as stated by Gaskell (Gaskell & Dumay, 2003b). In order to find a delay in the response of the detection of the pause, it is necessary for the word form to be lexicalized but not its meaning. In other words, "*appet*" will compete with "*apple*" whether we know the meaning of "*appet*" or not.

Our hypothesis is that the lexicalization process and consequent appearance of competition will be present for the words in the remote condition, which had undergone two nights of consolidation, but not for the words of the recent condition, learned the same day. Hence we expect to find the same pattern in L2 as in L1.

1.4 Semanticization - primed lexical decision task

As for the word-form level, also for the meaning-level, one of the strongest pieces of evidence that can be found for the semanticization of a word is its interaction with semantically related words or concepts. Two different tasks has been used in the literature to investigate this interaction, the picture-word interaction task (Lupker, 1979) and the primed lexical decision task (Verhoeven et al., 2015). Given our interest for the interaction between trained novel words and semantically related words we chose the primed lexical decision task, since the picture-word interaction task investigates the mutual influence between words and semantically related pictures, hence, without word-form involvement but instead focusing into the interaction directly at the conceptual level. Primed lexical decision task on the other hand reflects the interaction between concepts engaging the word-forms.

In the primed lexical decision task, participants have to decide if the target is an existing word or not. The target word can be a string of letters presented visually or presented as an auditory stimulus. When the target is preceded by a semantically related word, the prime, the RTs to the target words are faster than the RTs of the response when the prime word is semantically unrelated. This priming effect is caused by the activation of a semantically related network of words in the mental lexicon, including the target word (Patterson et al., 2007).

As with lexical competition, this priming effect can be used as evidence of lexicalization. If "*appet*" primes the word "*bird*", we could assume that the word "*appet*" has been integrated into the semantic network, and hence, its meaning has been lexicalized. There are two main kinds of semantic priming: relational and associative. We have a semantic relation when two words have overlapping meaning (i.e. mouse-rat) and a semantic association when two words are repetitively experienced together (i.e. mouse-cheese). In the literature no particular difference has been found between the two types of semantic priming (Perea, Duñabeitia, & Carreiras, 2008) and we used the two kinds of relationship indiscriminately.

It is important to underline that priming can arise for word pairs that were pre-experimentally unrelated (e.g., mouse-puzzle), but co-occurred in a study list just prior to the priming task (McKoon & Ratcliff, 1979). If prime-target connections have just been experienced, priming can result from the episodic (not yet integrated) memory of prime-target connections, rather than connections in the semantic lexicon (McKoon & Ratcliff, 1979; Tamminen & Gaskell, 2013). Hence, we took care that the target words of the primed lexical decision task, never occurred earlier in the test in combination with the trained novel words, in order to avoid this possible confound and hence make it more likely to observe an actual semanticization effect. Also, semantic priming can occur when either the prime or the target is presented in visual or auditory modality (Holcomb & Anderson, 1993) with a stronger effect when the prime is presented in auditory modality (Holcomb & Neville, 1990). Semantic priming has already been shown to arise in a known foreign language and between languages (Perea et al., 2008).

It is important to underline a difference between the primed lexical decision task and the picture-word integration task on the one hand and the pause detection task on the other. The pause detection task is independent of the meaning: the lexicalization effect can arise regardless of the memorization of the meaning. In the first studies on the topic, the words were learned without any associated meaning but the interaction with other words occurred anyway (Gaskell & Dumay, 2003a). On the other hand, the primed lexical decision task and the picture-word integration task use the word form in order to prime related concepts, which are expressed by pictures or words. Hence, in these tasks is essential to be able to connect the word-form to its related meaning but not the vice-versa, to connect a meaning to its word-form as in a picture naming task, because the word form in these tasks is provided by the design of the task by itself (i.e. you hear or read the word which will interact with other concepts). Thus, with these tasks the more challenging aspects of the integration of the word-form are avoided.

In spite of that, our hypothesis is that the semanticization process and consequently the arise of priming, will be present only for the words of the Remote condition, which had undergone two nights of consolidation, but not for the words of the Recent condition, learned the same day. Hence we expect to find the same pattern in L2 as in L1.

1.5 Second language learning

Many researchers have investigated the L2 learning process. In spite of that, only a few studies have investigated the word learning process in L2 in the terms discussed here. We wanted to examine if the word learning process is the same between L1 and L2 and on top of that if any individual difference in vocabulary size, length of stay and proficiency would influence in any manner in this process. Overall, previous literature seems to agree that the word learning process follows the same mechanism in L1 and L2 (Pajak, Creel, & C., Levy, 2016; Ullman, 2001) but there has been little work on the individual differences in this process.

About the role of proficiency, some researchers show an interaction between the level of proficiency and integration effects (Guasch, Sánchez-Casas, Ferré, & García-Albea, 2008; J.F. Kroll & Stewart, 1994; Judith F. Kroll & Sholl, 1992) but the literature on the issue is not always coherent. Studies with a translation recognition task sometimes found that cross-language lexicalization and semanticization effects vary as a function of proficiency in L2 (Adrienne Talamas, Judith F. Kroll, & Robert Dufour, 1999; Sunderman & Kroll, 2006) while other times it has been found integration effects with proficient bilinguals but only lexicalization effects with non-proficient bilinguals (Ferré, Sánchez-Casas, & Guasch, 2006). However, these studies used different tasks than ours and did not focus on the mnemonic aspect of learning a word, since all the words used were already known by the participants. In addition, they looked into cross-language interaction while in this research we focused on the interaction within L2.

Nevertheless, our hypothesis is that proficiency level will have an influence on the memorization, consolidation, and integration processes. Thus it will correlate with the size of the effect of both priming and competition effects at the behavioral level and also with the pattern of activation of the hippocampus and the pMTG. Previous literature has shown that the hippocampal activation is a good predictor to differentiate good and poor learners (Breitenstein et al., 2005), and since higher proficiency participants should show a better integration, and we tie integration to consolidation, with the latter with a shift from hippocampus to pMTG activation, it is reasonable to expect such an outcome. Specifically, we expect that proficiency will correlate with pMTG activation and anti-correlate with hippocampal activation in the remote condition. Also, we expect to find a positive correlation between the proficiency level and the performance in the free memory recall performed at the end of the training phase in both days, as for the learning curve of the training phase and the performances of the picture naming tasks conducted during training. We choose to investigate these tasks because are the more challenging and the best representation of the actual learning of a word-form and the connection to its concept. On the other hand, it could be that word learning is not influenced by the proficiency. If this is the case, the proficiency measurement will not correlate with any result in the training phase and neither in the testing phase. Thus, word learning would be independent by our proficiency in that language, and hence, an ability which is independent from which language we are learning the word. Hence, word learning would be a cross-linguistic ability which does not vary between languages.

As far as we know, no specific research has been conducted on the influence of length of stay in a foreign country on word learning. However, length of stay seems to be beneficial for second language acquisition (DuFon, 2004; Llanes & Muñoz, 2009) even for a short stay (Llanes & Muñoz, 2009). Given that the length of stay should increase the proficiency level, the two measurements should highly correlate with each other and therefore have the same influence on the processes we are interested in.

With the respect to vocabulary size, previous experiments (James, Gaskell, Weighall, & Henderson, 2017; van Goch, Verhoeven, & McQueen, 2017) suggests that "the rich get richer" (Matthew effect), hence that a larger vocabulary size should help word learning. Thus, as for the previous covariates, we predict a beneficial influence of vocabulary size on the processes investigated here.

In order to control for these variables, we, therefore, collected the IELTS score, the number of months spent in an English speaking country, and the number of unknown words among the words used in the experiment in a pre-test.

1.6 L1 vs. L2, the provision of word form and the time scale.

At this point it is important to underline a possible difference between L1 and L2 in word learning. When we learn a new word in L1, we also learn a new concept to associate with the word, while when learning a new word in L2, the concept it is already integrated. For example, an adult native English speaker could learn the word "*omniscience*", which means "*the quality of knowing everything*". Even if the native speaker could already think that concept before knowing the word, he or she never knew that it was possible to express it in one single word, hence, it has to link the new conceptual pattern with a new word-form. On the other hand, when we learn a new word in L2, often the concept that we have to link to the new word-form is already present, as the conceptual links related to that concept. For example, when an Italian native speaker has to learn the word "*chair*", the concept expressed by the word and its semantic relations (i.e object, table, furniture where you sit etc.) are already integrated. However, this should not be the case of our experiment since in our training, we explicitly told the participants that the concept related to the word they were going to learn, were specific concept. For example, that "*appet*" is that specific kind of bird, and it does not mean just "*bird*". Participants should link "*appet*" to the already existing concept of "*bird*" and hence, "*appet*" should prime "*bird*" in the primed lexical decision task.

Also, if the participants are able to connect the specific concept to its word-form, they should correctly perform the picture word association task in the scanner, where they have to link a new picture to its word-form, which was provided among the four options. The fact that the picture is a new one and they have to correctly link it to its word-form, is informative if they are able to extract its characteristic features and recognize it, without relying on the picture by itself.

In addition, in some research, semantic priming have been found immediately after training, even when the possible confound of episodic semantic association was taken out (Bordag et al., 2015; Kapnoula & McMurray, 2016; Kapnoula et al., 2015; Lindsay & Gaskell, 2013.; Oppenheim, 2018; Szmalec et al., 2012). As well, some experiments found immediate lexicalization after training (Kapnoula, Packard, Gupta, & McMurray, 2015; Szmalec, Page, & Duyck, 2012; Lindsay, S., & Gaskell, M. G. 2013) but an important reflection about the tasks used to test lexicalization and semanticization must be done. A key factor for the immediate occurrence of integration effects could be whether the newly learned word-form is provided or not during the task. For the investigation of the semanticization process is impossible to avoid the usage of the newly learned word form, otherwise we would look to pure conceptual interaction unable to link the discussion to the linguistic aspects of the semanticizations. Hence, to investigate semanticization we must provide the word-form (i.e. appet). On the other hand, to investigate lexicalization this is not mandatory and we can avoid it as in the lexical competition – pause detection task. In the lexical competition - pause detection task, the investigation is performed on the base-word (i.e. apple) of the newly learned word (i.e. appet).

Another way to investigate lexicalization is the usage of the visual word paradigm (Kapnoula & McMurray, 2016; Kapnoula et al., 2015). In this task, participants have to look to the image which represents the existing word they just heard. The listened word can go through different manipulations, the important one for the purpose of our discussion, is when the listened word is created using some parts of a previously newly learned word. For example, the word "*job*" is created using the phoneme "*b*" and the phonemes "*jo*" extracted from the pseudo-word "*jod*" previously learned. Hence, we have two possible conditions: the normal version of the word "*job*" and the artificially created word "*jo_b*" with the phonemes "*jo*" from the previously learned pseudo-word "*jod*". When "*jod*" has been lexicalized it is supposed to compete with "*job*" retarding the fixation of the correct target picture, similarly to the competition which arises in the lexical competition– pause detection task. Even if the two tasks could seems to investigate the same effect, to provide (part of the) the word form as in the visual word paradigm, could have significant effect on the processes involved. If the word-form of the newly learned word is provided, as in the visual word paradigm, it could be that the memorization and neither the integration of the word is needed, since the participants do not need to recall it from the memory. From the literature we also know that semantic association (i.e. mouse - cheese) can occur directly after training and could represent an episodic hippocampal connection rather than connection in the mental lexicon (Breitenstein, Caterina Zwitserlood et al., 1989; Dagenbach et al., 2007). Similar to this problem, with semantic associations which can occurred directly after training and representing an episodic hippocampal connection rather than connection in the mental lexicon, it could be that the competition that we see immediately after training with the visual word paradigm, is not based on an actual connection of the word form in the mental lexicon but rather a competition with a fresh episodic hippocampal memory triggered by the provision of the word. In order to make more robust this reasoning we can observe the results of two different researches which found lexicalization without sleep. On the one hand, there is research which found lexicalization immediately after the training (Kapnoula & McMurray, 2016; Kapnoula et al., 2015) with the visual word paradigm, on the other hand there is research which found lexicalization without sleep but not immediately after training but only after 12 hours (without sleep) (Lindsay & Gaskell, 2013.; Szmalec et al., 2012). Hence, if we test with the visual word paradigm (providing the word) we can see competition immediately after training, and if we test with the pause detection at least twelve hours are needed for the competition to arise. This comparison seems to agree with our reflection on the possibility that some competition effect found in previous literature do not represent an actual insertion of the word in the mental lexicon but rather a competition with an episodic trace of the word-form trigger by the provision of the word. This is the reason why we expect to find lexicalization and semanticization effects only in the Remote condition.

Thus, to provide the word-form during the task seems to play an important role in the processes discussed in this research. Nevertheless, we decided to provide the word-form using the picture-word association task in the scanner. We took this decision because a picture naming task would have spoiled the imaging data due to the movements done fort the articulation. Also, a silent picture naming task would not have been informative of the correct retrieval of the word-form leaving us unable to remove the incorrect trials.

1.7 Hypothesis

1 Neural data will show a shift in the retrieval network from the hippocampus to the pMTG with hippocampal activation in the Recent condition and pMTG activation in the Remote condition.

2 Lexicalization will follow the same pattern as in L1, with slower RTs only for the Remote condition in the lexical competition task – pause detection. We also predict that consolidation process is beneficial for the occurrence of integration effects at the behavioral level.

3 Semanticization will follow the same pattern as in L1, with faster RTs only for the Remote condition in the primed lexical decision task. . We also predict that consolidation process is beneficial for the occurrence of integration effects at the behavioral level.

4 The measurements of the individual differences, IELTs, length of stay, vocabulary size, will correlate with the performance in the free memory recall and picture naming tasks of the training phase. These measurements will also correlate with pMTG activation and anti-correlate with Hippocampal activation in the remote condition.

2 Methods

2.1 Participants

Fifty-two right-handed native Italians were recruited to take part in the study, 22 males, average age 25 years, range 18 - 41. All participants had no history of neurological or language-related disorders and reported having normal or corrected-to-normal vision and hearing. One participant was excluded from the analysis because she was not able to conduct the experiment in the scanner due to an attack of claustrophobia. Three more participants were excluded given their slow RTs (above the threshold of 2.5 standard deviations above the group mean) in one or more of the primary behavioral tasks. Thus the data that were analyzed were based on 48 participants.

2.2 Material

2.2.1 Words and pseudo-words

One hundred and twenty pseudo-English words were created and divided into three different sets of 40 words each. The novel words in each set were balanced for orthographic length (range 4 – 11 letter) and phonological length (range 3 – 10 phonemes), for the number of syllables (range 2 - 5) and for frequency of occurrence (range 1.1461 - 4.3457 log10 word frequency per million). These calculations were conducted with SubtlexUS (Brysbaert & New, 2009) and CLEARPOND (Marian, Bartolotti, Chabal, & Shook, 2012).

The pseudo-words were created from existing English words (hereafter called base words) with the software Wuggy (Keuleers & Brysbaert, 2010) a multilingual pseudo-word generator. For each of the base words, one or two phonemes of the last syllable of the word were changed, in order to create words which do not exist in English or Italian (e.g. apple - appet). If the word was monosyllabic, two phonemes were added at the end of the word (e.g. cliff - cliffon).

The base words from which the pseudo-words were created were chosen to have the following features: all the words were concrete nouns in order to have words which do not differ in semantic and syntactic properties and which have broadly similar patterns of activation in the brain (Costa et al., 1999). In addition, there were no compound words in order to avoid the use of strategies (i.e. remember stepfather as step plus father), and without correspondence in Italian (i.e. cognate pairs such as cathedral/cattedrale) to force the subjects to use only their English lexicon.

Finally, all the words had small phonological neighborhoods (80% with less than three neighbors , min 0 max 6) and small orthographic neighborhoods (86% with less than three neighbors, min 0 max 7). In this way, we enhanced the competition effect in the lexical competition task.

In addition, 240 common words were selected and pairs of them were matched with each of the 120 pseudo-words. One of the two common words in each pair was semantically related to the concept expressed by the pseudo-word (see section 2.2.3 for further details) while the other was not semantically related to it. A further 240 pseudo-words were created with Wuggy from the 240 related and unrelated words previously described. Those words were used in the primed lexical decision task as non-existing words. Since the non-existing words were created from the related and unrelated words, they were necessarily matched on phonologic, syllabic and orthographic length to the related and unrelated words. Another 40 common words were selected and used as fillers in the lexical competition task (see Appendix A for further details).

2.2.3 Pictures

One hundred and twenty unfamiliar objects were selected and two photos depicting the objects were selected from the internet. The pictures were found on the internet using Google and any letters or words present in the pictures were blurred. The images selected were checked by 12 people who did not take part in the actual experiment. They were asked if the object looked meaningful (i.e. if the object was recognizable), and whether the two pictures could be considered to be the same object. . Pictures were chosen only if they met these criteria.

The 120 novel words were matched with two different pictures of the same object for a total of 240 pictures. The two pictures could differ in the angle of the photo, the color of the object or other features (see Fig 3). One set of pictures was always used for the training phase and the other one for the task conducted in the fMRI scanner.

All pictures represented meaningful objects and, were selected from different semantic categories (tool, animal, furniture etc.) which were equally distributed in the three different sets of novel words. A given novel-word was always matched with the same picture.

Fig. 3 An example of pictures used in the training and test phase.

2.2.4 Speech recordings and speech editing

All 120 base words, the 120 novel words derived from the base words, and the 40 filler words for the lexical competition task were recorded three times with a microphone and pronounced by a young female native speaker of American – English. The best of the three recordings was selected. In order to create materials for the pause detection task a silent pause of 200 ms was inserted in the base words and in the filler words immediately before the point of divergence between the base word and the pseudo-word (i.e. app_le vs. app-et) using the program AUDACITY. Fade in and fade out effects were applied in order to prevent any clipping sound generated by the manipulation.

2.3 Procedure

Before the beginning of the main experiment, participants completed the pre-test. If they were eligible for the experiment they were scheduled for two days of testing (see section 2.3.1 for further details).

The subjects were assigned to one of the 6 possible groups. Each group had one of the three different sets of novel words used for the first day, one set for the second day and one set was used as New condition for the task in the scanner (see section 2.3.3.2). On the first day, participants conducted the training phase with one of the three set of novel words (see Table 1 for an overview of the procedure). After that, a free memory recall test was performed. Two days after the first day of testing, hence with two nights allowing for sleep-related memory consolidation, the subject repeated the training phase with one of the other two sets of novel words and then the free memory recall task.

After that, the task in the fMRI, the lexical competition task, and the primed lexical decision task were performed in that order, with the latter two tasks performed by a computer in a behavioural booth (i.e. not in the scanner). At the end of the testing phase participants were informed of the purpose of the experiment and that the learned novel words were pseudo-words (if they did not figure this out during the experiment).

2.3.1 Pre-test

Before the beginning of the experiment, subjects conducted a preliminary screening in order to check that they were eligible for the experiment. Participants had to be Italian native speakers, without metal part in the body (for safety during the scanner session) and without claustrophobia and right handed. Given the high lateralization of language processing in the brain and the reverse activation of left handed subjects this could have created problem in the fMRI contrasts.

In addition, knowledge of all English words used in the experiment was tested in order to confirm which of the words were unknown to the participants. This was done because, obviously, in order for lexical competition to arise, between the novel words and the base words, it is necessary that the participant already knew the base words. The same can be argued for the target words used in the primed lexical decision task. The concept of the novel words will prime their related targets words, only if participants know the meaning of the target words.

Table 1

Day 0	Day 1		Day 2	
Pre-test	Training:	1 Familiarization	Training:	1 Familiarization
		2 Multiple Choice – Pictures – Oral repetition		2 Multiple Choice – Pictures – Oral repetition
		3 Multiple Choice – Words – Written repetition		3 Multiple Choice – Words – Written repetition
		4 Words Completion – Oral		4 Words Completion – Oral
		5 Multiple Choice – Pictures – Written repetition		5 Multiple Choice – Pictures – Written repetition
		6 Multiple Choice – Words – Oral repetition		6 Multiple Choice – Words – Oral repetition
		7 Picture Naming – Oral		7 Picture Naming – Oral
		8 Multiple Choice – Pictures – Oral repetition		8 Multiple Choice – Pictures – Oral repetition
		9 Picture Naming – Oral		9 Picture Naming – Oral
		10 Familiarization		10 Familiarization
	Free memory recall		Free memory recall	
			Test	1 fMRI – Recognition memory task
				2 Lexical competition - pause detection task
				3 Primed lexical decision task

Table 1 – An overview of the whole procedure, including the two days of training and the testing phase with the related tasks.

The number of total unknown words was used as a proxy measurement of the vocabulary size of the subjects. The pre-test also included collections of the participant's proficiency in English using (when available) the IELTS score (when not available, other tests, such as TOEFL, TOEIC or Cambridge CELA were converted to the equivalent IELTS score). We choose the IELTS as measurement of proficiency since is widely considered a reliable evaluation of the level of English (Charge & Taylor, 1997). The length of stay in English speaking countries was noted. Given the generally high level of English proficiency in the sample, the fact that none of them spoke Dutch, and the international English-speaking environment they all lived in, the Netherlands was considered to be an English speaking country.

2.3.2 Training phase

At the beginning of the training phase, instruction about the purpose of the experiment and the procedure was given. Then ten different tasks were conducted in order to help the subjects memorize as many novel-word picture associations as possible (see Fig. 4 for an example). They were trained on the spelling of the words, the sound of the words and their pronunciation (i.e. reading, listening and speaking). For each trial they were always

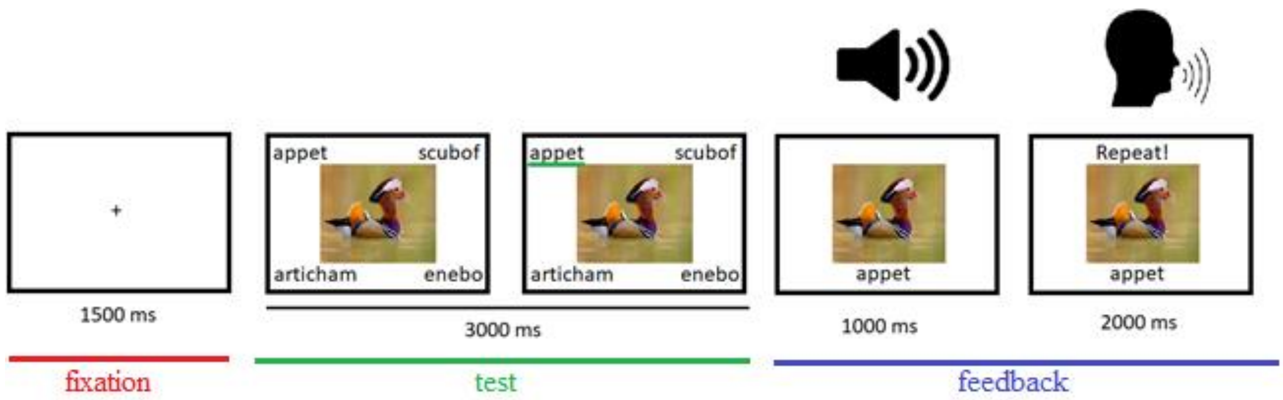

Fig 4 – An example of one of the tasks used in the training phase (multiple choice – words – oral repetition).

informed if the given answer was correct or not and, at the end of each task, the total number of correct answers was shown. All tasks were presented using the software Presentation (www.neurobs.com) and used a button press device and a computer keyboard.

2.3.2.1 Familiarization

The training phase started and ended with a familiarization task. This was done at the beginning of the training phase, to give initial exposure to all the novel words and their associated pictures and, at the end, as the last control of participant's knowledge of the novel word-picture pairing. During this task, a cross appeared in the middle of the screen for 1 second. Then the picture of the object with the associated novel word appeared. After 500 ms, the word was presented auditorily through the speakers. At the offset of the novel word, the word "Repeat!" appeared on the screen and the subjects were instructed to repeat the word out aloud within 3 seconds. All the 40 novel word-picture pairs were presented one time each in a random order.

2.3.2.2 Multiple Choice – Pictures – Oral repetition

This task was conducted as the second and again as the eighth task of the training phase. It was positioned at the beginning and end of the procedure in order to have a measure of the learning curve of the subjects. During this task, a cross appeared at the center of the screen for 1 second. Then a pseudo-word (already presented in the familiarization task) appeared in the center of the screen with four different images (which had already been presented during the familiarization task) were displayed in the four corners of the screen. One of them was the correct image (i.e. was paired with the novel word). After 100 ms the sound of the word was presented through the speakers. At acoustic offset, the subjects had to correctly associate the word with its image by selecting one of the four buttons of the pad, each of them representing one of the four corners. A colored bar under the picture of the given answer, informed the subject if the association was correct (green bar) or incorrect (red bar). Participants were instructed to wait for their response until they had heard the full word. All the answers given before the end of the sound were ignored. After 3 seconds, if the subject did not answer, the pseudo-word on the screen disappeared and the four pictures were substituted with digits 1,2,3,4, indicating the corresponding button press. When a button response was made, a colored bar appeared below the chosen picture digit. Also here the colored bar indicated the correctness of the answer. After the answer, the correct picture appeared at the center of the screen with the novel word under it. After 100 ms the sound of the word was presented. At the offset of the sound the word "Repeat!" appeared on the top of the image and the participants were instructed to repeat the word within 2 seconds. This was done for all 40 trained picture-word associations, one time each in a random order.

2.3.2.3 Multiple Choice – Words – Written repetition

The third task was again a multiple choice task. At the beginning, for 1500 ms a cross appears in the middle of the screen. Then a picture appeared in the middle of the screen with four novel words from the trained list in the four corners with one of them being the correct option. As in the previous task, a colored bar under the given answer notified the participants if the answer was correct (green) or not (red). After 3 seconds, if the subject did not answer, the options and the picture on the screen disappeared. The four pseudo-words were substituted with 1,2,3,4, as before, those numbers indicating the corresponding button press. Also here, a colored bar indicated the correctness of the answer. After the response, the correct picture and novel word appeared at the center of the screen and after 100 ms the auditory form of the word was presented. At the acoustic offset of the word "Write!" appeared on the top of the image and the participants had to write the word they had heard. When the subject pressed ENTER after he or she wrote the pseudo-word, the program moved to the subsequent trial. This was done for all 40 trained picture-word associations, one time each in a random order.

2.3.2.4 Words Completion – Oral

In the fourth task, for each trial, the participants were presented with a picture from the trained set. And their task was to correctly orally name the picture given the first letter as a hint. At the beginning of the trial for 1 second a cross appeared in the middle of the screen, then one of the images of the set was presented in the middle of the screen. On top of the picture the phrase "Name it!" was presented and under the picture, the initial letter of the corresponding novel word was shown. Participants had 4 seconds to correctly orally name the picture. After 4 sec the picture and the letter disappeared and the phrase "Press the space-bar to have feedback". After the response, the picture and the correct novel word under it appeared on the screen. After 100 ms the pronunciation of the word was presented through the speakers. At the offset of the spoken novel word the word "Repeat!" was shown on top of the picture. The subjects had 2 seconds to correctly repeat the name of the picture, then the program continued to the next trial. This was done for all 40 trained picture-word associations, one time each in a random order.

2.3.2.5 Multiple Choice – Pictures – Written repetition

The fifth task was very similar to the second and eight tasks. Here the subjects had to correctly associate a novel word presented in the middle of the screen with one of the four pictures shown in the four corners. The stimulus presentation until this point was the same as the one described in Multiple Pictures. The difference was in the format used for the repetition after the feedback, this time the participants had to write the correct novel word instead of orally repeat it. Once the answer was given, the correct picture was presented in the middle of the screen and after 100 ms the sound of the correct pseudo-word was presented. At the offset of the sound, the word “Write!” appeared on the top of the picture. After the button ENTER was pressed, the program moved to the next trial. This was done for all 40 trained picture-word associations, one time each in a random order.

2.3.2.6 Multiple Choice – Words – Oral repetition

The sixth task was very similar to the one previously described as multiple words – write. The stimuli presentation was the same as previously described, the difference was in the format of the repetition after the feedback, this time the subjects had to orally repeat instead of writing the correct answer. After the answer was given, the correct picture with the correct novel word under it appears on the screen. After 100 ms the sound of the word was presented and at the offset the word “Repeat!” appeared on the top of the picture. The participants had 2 seconds to repeat the word. After that the program moved to the next trial. This was done for all 40 trained picture-word associations, one time each in a random order.

2.3.2.7 Picture Naming – Oral

The seventh task was a picture naming task. In the beginning, a cross appeared for 1 second at the center of the screen, then a picture with the text “Name it!” on top of it appeared on the screen. The subject had 3 seconds to orally name the picture. After that, the picture with the correct name appears on the screen and after 100 ms the sound of the pronunciation of the correct novel words was presented. After that the word “Repeat!” appeared on the screen and the participant had 2 seconds to correctly repeat the novel word. This was done for all 40 trained picture-word associations, one time each in a random order.

2.3.2.8 Picture Naming – Written

The ninth task is very similar to the seventh one but this time the subject had to name the picture in writing and orally repeat the correct novel word after the feedback. At the beginning a cross appeared in the center of the screen for one second, then a picture with “Write! Press ENTER to have feedback” on top of it was shown. After the subject pressed the button ENTER, the program moved to the feedback phase. The picture with the correct name was shown under it, after 100 ms the pronunciation of the correct novel word was presented. At the offset of the sound the word “Repeat!” appeared on the screen and the subject had 2 seconds to correctly repeat the novel word. This was done for all 40 trained picture-word associations, one time each in a random order.

2.3.3 Testing phase

After the training phase on each day, the subject performed a free memory recall test. After that, but only on the second day of testing, the main testing phase was conducted.

2.3.3.1 Free memory recall

At the end of the training procedure the subjects had 5 minutes to type all the trained novel words he or she could remember.

2.3.3.2 fMRI – Recognition memory task (picture – word association)

For this task, a second set of pictures was used in order to avoid the usage of episodic memory and force the subject to link the concept expressed by the picture, and not the picture itself, to the matching novel word.

At the beginning of the trial, a black cross appeared for a jittered inter-trial interval (ITI) of 1 to 7seconds. After that, the cross turned blue for one second, prompting the participant that the trial would be presented. One picture was presented in the middle of the screen with four different options in the four corners of the screens. The picture could be from the following three conditions: the picture was trained on the same day (recent condition), studied two days before (remote condition), or new (new condition). One of the four response options was the correct novel word, one was an incorrect novel word from the same training set, the third was one novel word taken from the other training set and the fourth was "new object". The positions of the four options were randomized at every trial. In the new condition the correct novel word was substituted with a word from the untrained set and the correct answer was "new object".

Participants had 2500 ms to answer, otherwise, the trial was considered as a wrong answer. After the answer, a black bar appeared under the chosen option for 500 ms, but no corrective feedback was given, hence, the subject did not know if the answer was correct or not. This procedure was repeated for all 120 pictures one time each in random order.

2.3.3.3 Lexical competition - pause detection task

In this task, the participant had to identify as fast as possible if a pause was present or not in spoken words. This task was composed of 160 trials, 40 of them were with the base words from which the remote novel words were created, 40 with the base words from which the recent novel words had been created, 40 with the base words from which the novel words of the third unused set (new condition) had been created, and 40 with filler words (filler condition). Half of the trials were with a pause inserted in the base word and half of them without a pause. The presence of the pause was randomized such that half of the trials (80) was with a pause.

At the beginning of the trial a cross appeared in the middle of the screen, then it disappeared and the spoken word was played at the same time three response options appeared on the screen: "Pause" on the left, "no pause" on the right and in the middle-down part of the screen the option "Spacebar - No answer". The participants used the button pad for the first two options and the space bar on the keyboard for the third. The participants could answer only after the onset of the pause, all answers given before were ignored. If the pause was not present, the onset of the pause in the corresponding record of the same word with the pause was used as time delimiter for ignoring the answers given before. After the answer, a black bar appears under the chosen option for 500 ms, but no feedback was given. If the subject did not reply in less than 3 seconds after the offset of the sound, the answer was counted as a miss.

2.3.3.4 Primed lexical decision task

In this task, the participants heard a novel word from the trained session, and then saw a word on the screen. Their task was to decide whether the word on the screen exist or not in English. This task was composed of 320 trials, half of them were with the remote novel words as the prime and the other half with recent novel words as primes. Each prime word appeared 4 times during the task, one time with a related word, one time with an unrelated word, and two times with two different non-existing words as target word. Hence there were six different conditions, the priming novel words could be consolidated or unconsolidated, and the target words could be related unrelated or not existing. Prime word order was randomized and every sequence of 4 trials contains 1 related, 1 unrelated and 2 pseudo-words as target word in random order. After the trials 80, 160 and 240 there was a break in order to give to the subject time to rest and prepare for the subsequent part.

At the beginning of the trial a cross appeared in the center of the screen for 1 second, then it disappeared and the prime novel word was presented auditorily. Participants were instructed to think of the meaning of the presented word. At acoustic offset of the prime, a target word appeared at the center of the screen together with the response options. On the left part of the screen the option “Real word” was presented, on the right part of the screen the option “No word” and in the middle-down part of the screen the option “Spacebar - No answer”. The participants used the button pad for the first two options and the keyboard for the third. After the answer, a black bar appeared under the chosen option for 500 ms, but no corrective feedback was given, hence, the subject did not know if the answer was correct or not. If the subject did not reply within 4 seconds after target word onset, the answer was categorized as a miss.

2.3.3.5 fMRI – technical details

fMRI data were recorded in a 3 T Prisma scanner (Siemens Healthcare, Erlangen, Germany) using a 32-channel head coil. For functional images, we used a multiband sequence with the following parameters: repetition time (TR): 1000 ms, echo time: TE 34 ms, 66 slices, ascending slice order, 2.0 mm slice thickness, inplane field of view (FOV): 210 x 210 mm, flip angle: 60°. Slices were angulated in an oblique axial manner to reach whole-brain coverage. In addition, an inverted EPI with the same parameters was collected. Also, for field-map images, we used a multiband sequence with the following parameters: TR: 620,0 ms, echo time: TE 1 of 4.70 ms and TE 2 of 7.16 ms, 66 slices, ascending slice order, 2.0 mm slice thickness, inplane field of view (FOV): 210 x 210 mm, flip angle: 60°. T1-weighted anatomical scans at 1 mm isotropic resolution were acquired with TR 2300 ms, TE 3.03 ms, flip angle 8°, and FOV 256x256 mm.

Image pre-processing and statistical analysis was performed using SPM12. The multi-band sequence was with a acceleration factor of 6 and the first ten volumes of each participant's functional scan were discarded to allow for T1 equilibration. Field-map images, T1 structural images and functional images were then converted from DICOM files to nifti files to allow the use of standard fMRI preprocessing tools in SPM12. Successively, the field-map deformation was calculated and the functional images were realigned and un-warped. After that, the subject-mean image of the functional run were co-registered to the corresponding structural MRI and applied to all functional scans. Consecutively, T1 structural images were segmented and the functional images were first normalized and then smoothed. The fMRI data were analyzed statistically using a general linear model (GLM) and statistical parametric mapping. Four explanatory variables were included in the model for each session: correct trials for remote condition, recent condition, new condition (untrained) and incorrect trials pooled across different conditions. The first three variables contained only correct trials, the last one only incorrect trials. These explanatory variables were temporally convolved with the canonical Hemodynamic Response Function (HRF) provided by SPM12. Each event was time-locked to the onset of the picture. The design matrix included the six head motion regressors (three translations, three rotations). A high pass filter was implemented using a cut-off period of 128 s to remove low-frequency effects from the time series. .

Two different regions of interest (ROI) masks were created using MARSBAR an SPM12 toolbox. One of the masks covered the anatomical area of the hippocampus defined by the HIP AAL template (Tzourio-Mazoyer et al., 2002) the other one of the pMTG area. The latter one was created combining two different masks, an anatomical mask of the whole MTG defined by the T2 AAL template (Tzourio-Mazoyer et al., 2002) and a sphere of 2 cm diameter with the center in -58 -60 0 (MNI coordinates). The coordinates were taken from the peak of activation in the left pMTG in Takashima study (Takashima et al., 2014).

3 Results

3.1 Training

Day 1	Absolute	Percentage	Day 2	Absolute	Percentage
2 Multiple Choice – Pictures – Oral repetition	13.6	34%	2 Multiple Choice – Pictures – Oral repetition	12.6	30%
3 Multiple Choice – Words – Written repetition	7.42	18.5%	3 Multiple Choice – Words – Written repetition	6.35	16%
4 Words Completion – Oral	23.24	58.1%	4 Words Completion – Oral	19.94	49.8%
5 Multiple Choice – Pictures – Written repetition	7.42	18.5%	5 Multiple Choice – Pictures – Written repetition	1.06	3%
6 Multiple Choice – Words – Oral repetition	1.7	4%	6 Multiple Choice – Words – Oral repetition	1.39	3.5%
7 Picture Naming – Oral	15	37.5%	7 Picture Naming – Oral	12.73	31.8%
8 Multiple Choice – Pictures – Oral repetition	0.75	2%	8 Multiple Choice – Pictures – Oral repetition	0.6	1.5%
9 Picture Naming – Writing	15.75	39.4%	9 Picture Naming – Writing	14.85	37.1%
Free memory recall	18.62	46.6%	Free memory recall	18.37	45.9%

Table 2 – An overview of the error rates (absolute and percentage) in all the tasks in the training phase.

If a base word was unknown the related novel word was excluded from the whole analysis (on average 14.7%). Semantically related and unrelated words which were unknown and used in the primed lexical decision task were excluded from the analysis of this task as well (on average 2.7%). Subjects had the tendency to perform slightly better during the second day of training. With a two tailed paired t-test, we compared the two means, one for the first day and one for the second day, of all the means of errors in all tasks. They were significantly different ($p = 0.031$). This is probably due to the fact that on the second-day, participants already knew what to expect from the training (e.g. with the respect to tasks difficulty) and the tasks they had to perform. The participants consistently improved during the training with the errors from around 15 in the first task to almost no errors towards the end (except on the more difficult free recall task). Thus, the training procedure was successful in teaching the novel words to the participants.

3.1.2 Learning curve

The difference in performance between the two Multiple Choice – Pictures – Oral repetition tests accomplished within a training session is informative of the speed of learning of the participants. In order to verify if there was any influence of the individual differences on the speed of learning of the participants we used a bivariate correlation. We found that those 2 measurements of learning speed, one for the first day and one for the second, did not correlate with IELTS score (r = 0.813, p= 0.812), length of stay (r = 0.300, p= 0.875), and vocabulary size (r = 0.366, p= 0.135). No significant correlation was found, hence, those individual differences did not predict how fast and accurately a subject will be able to acquire new words.

3.1.3 Picture naming

The performances on the three different picture naming tasks on the training of both days was also correlated with the three measures of individual differences. In the Words Completion – Oral task, where the first letter was provided, vocabulary size was significantly anti-correlated with the rate of correct trials both on the first day (r = -0.301, p = 0.042) and the second day (r = -0.308, p = 0.037). The same held for the Picture Naming – Oral tasks on the first day (r = -0397, p = 0.005.) and the second day (r = -0.387 p = 0.007). Only vocabulary size was found to anti-correlated with the performances in these tasks. There is an anti-correlation, since the measure was taken the number of the unknown words, so the smaller the vocabulary score better the performance.

3.1.4 Free memory recall

The performances of the free memory recall task on both days was also correlated with the three measurements of individual differences. In opposition with our hypothesis, length of stay and proficiency did not correlate with the number of correctly remembered words. Interestingly, we found different results for the two days with the respect to vocabulary size. On the second day, no significant correlation was found, but on the first day we found an anti-correlation of the vocabulary size and the number of (correctly) remembered words (r = -309, p = 0.033). There is an anti-correlation, once again because the measure was taken of the number of unknown word of the set. Overall these data contradict our initial hypothesis of an influence of the individual differences in the memorization process except for vocabulary size which correlated with the number of remembered words but only on the first day. Probably the advantages of a bigger vocabulary can be easily overwhelmed with a better familiarization with the task difficulties and requests (i.e. what and how well should I learn).

3.2 Tests

3.2.1 fMRI – Behavioural

The average number of errors in the scanner was 6.52 (5.43%) with a minimum of 1 and maximum of 21. We tested whether the error distribution did not differ between the remote, recent, and new conditions with a one-way ANOVA. They did not (p= 0.816). We further tested whether the RTs were equal among the three conditions also, with a one-way ANOVA. Again, there was no significant effect (p = 0.496).

3.2.2 fMRI – Imaging

For the second level analysis, we compared the three different conditions, Remote , Recent and New. We expected pMTG activation when contrasting the Remote with the Recent condition. On the other hand, we supposed hippocampal activation for the opposite contrast, hence, Recent minus Remote. All the following statistics are based on the values at the cluster level (pFWE) and we performed the analysis using the whole brain images and the beta values extracted from the ROIs. We also created contrast with the images of the New condition.

In contradiction with our hypothesis, the whole brains contrast between Remote and Recent conditions (see Fig. 6) and the vice-versa (Recent - Remote) did not reveal any significant different pattern of activation in pMTG or in the hippocampus. However, the Remote - Recent contrast revealed a significant cluster of activation for the IFG (peak at -34 30 6, cluster size = 69, p = 0.026). When changing the significance level of the threshold of the cluster from 0.0001 to 0.005, pMTG (peak at -34 30 6, cluster size = 1698, p < 0.001) was found to be active in the Remote-Recent contrast but not in the opposite contrast suggesting the possibility that the consolidation process in L2 is slower than in L1

The contrast between Remote and New condition revealed the left supramarginal gyrus (peak at -52 -26 28, cluster size = 604, p = 0.000), left angular gyrus (peak at -42 -72 36, cluster size = 726, p = 0.000) right hippocampus (peak at 26 -24 -14, cluster size = 24566, p = 0.000) left superior frontal gyrus (peak at -18 26 46, cluster size = 124, p = 0.001), and right middle occipital gyrus (peak at 40 -74 38, cluster size = 120, p =0.001). In addition also the left MTG, the area we were interested in, was found to be active in this contrast (peak at -60 -28 -14, cluster size = 110, p = 0.002).

On the other hand, the contrast between Recent and New conditions showed the left precuneus (peak at -10 -62 24, cluster size = 29639, p = 0.000) left angular gyrus (peak at -42 -74 36, cluster size = 2155, p = 0.000) right middle occipital gyrus (peak at 40 -76 36, cluster size = 387, p = 0.000), and left middle frontal gyrus with two clusters (peak at 42 -36 22, cluster size = 132, p = 0.001 and peak at -28 30 50, cluster size = 414, p = 0.000). Also here, left MTG (peak at -54 -28 -16, cluster size = 187, p = 0.000) has been found significantly more active. Hence, both the contrast with the new condition revealed the left MTG, the left angular gyrus, and the right middle occipital gyrus as significantly more active.

Since the whole level analysis may not be sensitive enough to detect changes in activation in our area of interested (which were decided before the beginning of the experiment) we conducted a ROI analysis comparing the activation of the Hippocampus and pMTG. The beta values extracted within the ROI were investigated with two ANOVAs, with factors conditions Remote, Recent and New. One ANOVA was used to investigate the values of the Hippocampus and the other one for the values of the pMTG. Between the beta values of the pMTG: Remote (mean = 2,888, SD = 19638.34), Recent (mean = 7,524, SD = 16983.26), and New (mean = 3,498, SD = 17613.19); no significant difference was found (p = 0.793). In contrast the values of the Hippocampus: Remote (mean = 4,427, SD =15454.73), Recent (mean = 3,560, SD = 14561.63), and New (mean = -12,428, SD = 12033.94); were significantly different (p = 0.026). This difference was driven by the values of the hippocampus in the new condition (mean = -12.428) which were significantly different from the values of the other two conditions (mean Remote = 4.427, mean Recent = 3.560; see Fig. 5).

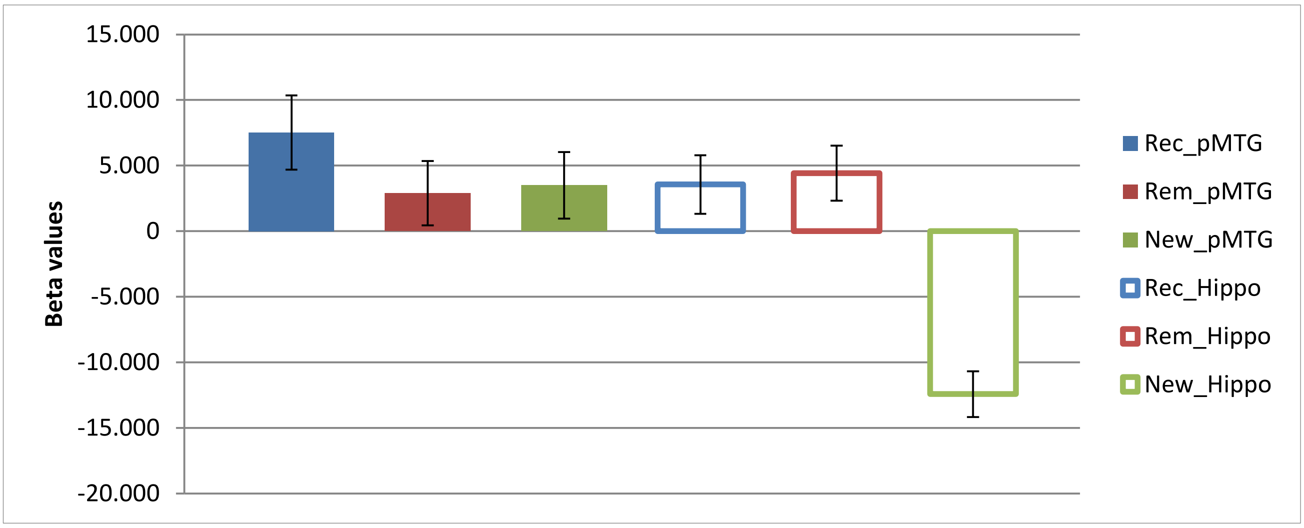

Fig 5 – The means of the beta values extracted from the pMTG and hippocampal ROI divided in the three conditions.

To investigate the role of the three covariates, IELTS, length of stay, vocabulary size in predicting the brain activation we created a model with the contrasts between Remote and Recent condition and the three covariates. We controlled if the three measurements correlated with the activation found in the contrast. We controlled the influence of the individual differences, one by one, in order to avoid an excessive spread of the explanatory power of the covariates, since they all highly correlate with each other. However, this investigation did not reveal any significant interaction. That is, none of the covariates predicted the patterns of brain activation we found, in contrast with our prediction

.

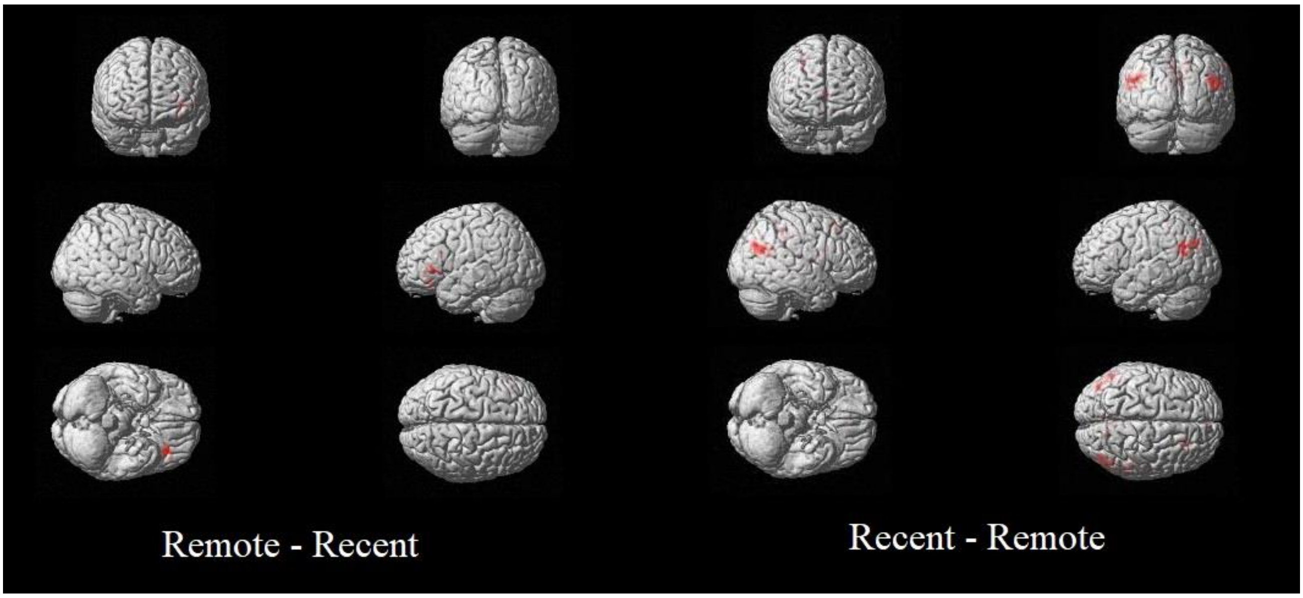

Fig 6 – The activation found in the two contrasts of interest: Remote minus Recent and Recent minus Remote.

3.2.3 Lexical competition - pause detection task

Responses above and beyond 2.5 standard deviations from the subject's mean were excluded from the analysis (10%). We hypothesized, in line with the previous literature in L1, that the Remote condition will have slower RTs compared to the Recent condition, since only the words of the Remote condition underwent two nights of sleep and are the only words which are expected to be lexicalized and hence, show the competition effect.

The results of this task are shown in Figure 7. With a repeated-measure design ANOVA, with two different factors: Remote vs. Recent, and Pause vs. Without Pause, we investigated if the base words of the remote condition have slower RTs compared to base words of the Recent condition. In line with previous literature, we did not find an effect regarding the presence or not of the pause ($p = 0.237$). As expected, the ANOVA revealed that the Remote condition (mean = 719.24 ms, SD = 167.25), was significantly slower ($p = 0.002$) than the Recent condition (mean = 686.58 ms, SD = 137.80), a result in line with the previous L1 literature. The effect arises regardless of the presence or not of the pause, and the RTs of the Remote condition are slower.

In order to investigate if the three covariates have any influence on the size of the effect (i.e. the difference of RTs between the Remote and Recent conditions), we correlated the size of the effect, with the three covariates, expecting correlation. The effect did not significantly correlate with IELTS score ($p = 0.224$) or length of stay ($p = 0.518$), and close to significantly anti-correlated with the vocabulary size ($r = -0.266$, $p = 0.068$). Since the vocabulary measure was the number of unknown words, this anti-correlation can be interpreted as the bigger the vocabulary, the bigger the effect, showing a trend in line with our hypothesis. However, the competition effect of lexicalization seems to be not influenced by the length of stay and proficiency, in opposition with our prediction.

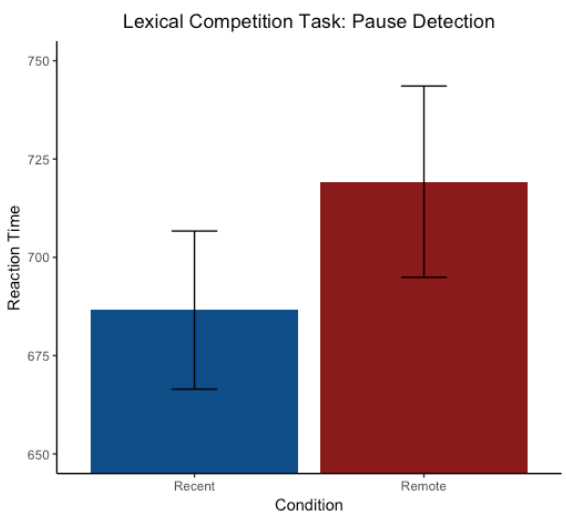

Fig. 7 – Mean of the RTs in the lexical competition – pause detection task for the two conditions, collapsing over the pause and without pause conditions.

3.2.4 Primed lexical decision task

Responses above and beyond 2.5 standard deviations from the subject's mean were excluded from the analysis (6%).We performed a repeated-measure design ANOVA, with two factors: Remote vs. Recent and Related vs. Unrelated.

We expected that only the words of the Remote condition would prime the target word when it is semantically related, speeding up the RTs of this condition. This is because only the words of the Remote condition underwent two nights of sleep and are the only words which are expected to be semanticized and hence, show the priming effect. We have therefore four different cases of interest: Remote-Related (mean = 720.89, SD = 128.99), Remote-Unrelated (mean = 753.14, SD = 147.64), Recent-Related (mean = 752.66, SD = 150.30), and Recent-Unrelated (mean = 753.69, SD = 176.16). The results of this task are shown in Figure 8.

In contrast with previous literature and our hypothesis, we found a significant effect for Related vs. Unrelated ($p = 0.033$), hence with a priming effect for Remote and Recent conditions merged together. However, the interaction between the two factors was found to be close to significance ($p = 0.054$), hence with the occurrence of the priming effect only with words of the Remote condition with the target words of the Related condition. We can thus argue that in both conditions a priming effect arose but numerically it is clear that this is mainly driven by the words of the Remote condition. A follow-up exploratory pairwise t tests showed a significance difference between the Related and Unrelated conditions in the Remote condition ($p < 0.001$) but not in the Recent ($p = 0.468$)

Also here, we investigate if the three covariates have any influence on the size of the effect (i.e. the difference of RTs between the Related and Unrelated conditions in the Remote condition), expecting correlation. Thus, we correlated the size of the effect, with the three covariates. Contrary to our prediction, the priming effect do not significantly correlate with IELTS score ($p = 0.101$), length of stay ($p = 0.834$), and the vocabulary size ($p = 0.098$).

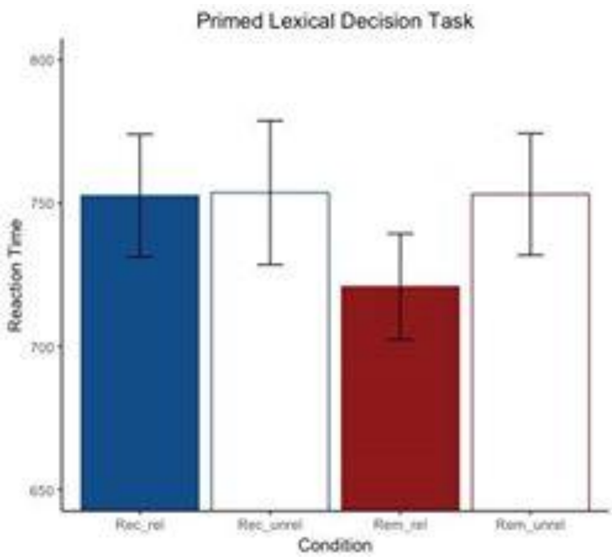

Fig. 8 – Mean of the RTs of the primed lexical decision task in the four conditions.

3.3 Summary

3.3.1 Behavioural results

The training was successful and accomplished the objective of making participants learn as many words as possible.

The lexical competition task – pause detection, as hypothesized, showed the same effect, with the same size and direction, as found in previous literature, with slower RTs for the Remote condition only. The pattern of lexicalization, hence, is the same as in L1.

On the other hand, the primed lexical competition task shown a priming effect regardless of the Remote and Recent condition. In spite of that, the interaction between the Remote condition and Related condition is close to significance, pointing towards a trend in line with the previous literature, as shown by a significant follow-up exploratory pairwise t tests. However, no strong conclusion can be drawn from the data about a similar pattern of semanticization between L1 and L2.

In contrast with our prediction, the covariates length of stay and proficiency did not show any correlation with the lexicalization, semanticization and consolidation processes. However, vocabulary size correlated with performance on the picture naming tasks, the free memory recall on day one and the size of the lexical competition effect, showing the expected correlation or trend.

The behavioural results of the task in the scanner shown a successful memorization of the learned word, hence, the task can be considered a reliable measurement of the memorization process.

3.3.2 Imaging results

The whole brain analysis did not shown the forecasted pattern of activation. The hippocampus was found to be active in the Recent condition, as predicted, but also in the Remote condition which is unexpected.

The same can be argue for the ROI analysis, with no significance difference of activation in the pMTG between the three conditions (Remote, Recent, and New) and a significantly different hippocampal activation only for the New condition but not for the other two.

Contrary to our prediction, the individual differences did not correlate with any pattern of activation in the brain.

4. Discussion

With this research we wanted to investigate: firstly if the integration of novel words in L2 is the same as in L1 and if the different aspects of integration, semanticization and lexicalization, have the same pattern; secondly if the consolidation process in the brain is beneficial for the appearance of these two effects and which are the brain areas involved in these processes; thirdly if proficiency in L2, vocabulary size in L2, and length of stay in an L2 speaking country, play a role in this process.

4.1 Behavioral

4.1.1 Lexicalization

As expected, for the lexicalization process we found the same pattern as in L1, with only the base words of the remote condition slowing down the response. Also the size of the competition effect, the difference of RTs between the Remote and Recent condition, is in line with previous L1 literature. We found a difference of 32 ms when in the literature usually vary between 20 and 40 ms (Cluff & Luce, 1990; Dumay, Frauenfelder, & Content, 2000; Goldinger, Luce, & Pisoni, 1989; McQueen, Norris, & Cutler, 1994.; Norris, McQueen, & Cutler, 1995; Vroomen & de Gelder, 1995). Given the high significance, the direction and the size of the effect in line to the previous literature as the effect, this results can be considered as validation of the lexicalization of the word-form of the word in the lexicon. Hence, we can argue that the lexicalization process of the base-word form of words in L2 follow the same pattern as in L1 from a behavioral perspective.

4.1.2 Semanticization

Our results on the semanticization process are not fully in line with previous literature in L1. We expected, in the Related condition, a priming effect only for the Remote condition, instead we found a priming effect in both Remote and Recent condition and a close to significance interaction between Remote and Related conditions. However, in both Recent and Remote conditions we found the priming effect, the difference of RTs between the Related and Unrelated conditions, in the expected direction. In the Remote condition, we found a priming effect of 31 ms with the size of the effect that usually varies between 20 and 40 ms in the literature (Patterson et al., 2007; Perea et al., 2008; Verhoeven et al., 2015), while for the Recent condition we found a priming effect with the size of 1 ms. Hence, semanticization process seems to be slightly faster than the lexicalization process but like lexicalization, to be stronger after a consolidation period (i.e. sleep). The results, even if not perfectly in line with our prediction, seem to point towards a similar pattern of learning as in L1, with a strong trend towards a priming effect only for the Remote condition.

4.1.3 Individual differences

Contrary to our prediction the covariates considered, proficiency in L2, and length of stay in an L2 speaking country, did not correlated with the integration process, the memory performance, and the brain activation. The differences showed by the subjects seem to be driven more by individual differences, for example with some talented low proficiency participants performing better than untalented high proficiency subjects. Some people appears to be better in learning new words (or languages) than others. On the other hand, vocabulary size has been found significantly correlated with some of the investigated measures.

4.1.3.1 Proficiency

This measure was taken using the score at the IELTS test (or similar tests translated to IELTS scale) which is a detailed and reliable assessment of the English skills of the participants (Charge & Taylor, 1997). Against our prediction, it did not correlate with priming effect and neither competition effect. Also it did not correlate with the memory performances (free memory recall and picture naming tasks) of the participants and neither pMTG or hippocampal activation. It could be the case that we fail to find significant correlations because all the participants were already proficient enough to minimize the differences between them. All participants had a score of 6 or more, hence, enough to hang a normal conversation in English, whereas the previous literature usually confronted low proficiency / beginner of L2 with high proficiency L2 speakers. It could be the case that there is a critical point in L2 acquisition after which the proficiency is not significant anymore in word learning. Before this critical point, learners of L2 have considerable difficulties in the word learning process, while after the crossing of this critical point, everybody, regardless of the proficiency, can easily deal with word learning. If this is the case, the proficiency would not linearly correlate with the processes investigated here as we expected. However, other researches seems to indicate that word-learning is equally challenging and not influenced by proficiency (Pajak et al., 2016.). Both in L1 and L2 learn similar words is equally difficult, hence, a better phoneme categorization (due to a higher proficiency) to understand the differences in L2 words, would not lead to a better memory performance. Our results are in line with this perspective with no influence of proficiency in L2 word learning. Thus, word learning appears to be a process which occurs in a similar vein in L1 and L2 and to be independent from the proficiency of the language.

4.1.3.2 Length of stay

This measure was used as a different measure of the proficiency. A longer stay in an English speaking country should enhance all the skills involved in learning English. For example, longer stay should improve writing and listening skills with a higher knowledge of the spelling and phonemes not present in the native language. We tested Italians who just arrived abroad, hence, with few months of stay, and Italians who spent more than one year in an English speaking country. As with the proficiency, overall better knowledge of English seems to not help the memorization integration and consolidation of new words with this process independent and not influenced by the amount of time spent in an English speaking country.

4.1.3.3 Vocabulary size

The number of unknown words is the only individual difference that we found significantly correlated with memory performances: the free memory recall and picture naming tasks, which is in line with previous literature (James, Gaskell, Weighall, & Henderson, 2017; van Goch, Verhoeven, & McQueen, 2017). However, it did not correlate with priming effect and neither competition effect. Also it did not correlate with pMTG or hippocampal activation of the participants. Hence, vocabulary size seems to help the memorization process but not the integration or the consolidation processes. It could be that having a larger vocabulary helps to memorize the word-form of a new word, which is the more challenging aspect of word learning, compared with the memorization of a concept as previously argued. All the words participants had to study were very similar to the base word, hence it could be that already having a similar word stored in our mental lexicon facilitate the memorization since the participants can refers to the already integrated word and then "change" the already know words in the newly learned word. This process became particularly beneficial when the word-form is not provided, as in the tasks which have been significantly correlated with the dictionary size. For example, when performing the free memory recall, the participants could use a strategy like: "I remember there was a word similar to *apple*, I remember only the end was different, it was *app_something*, it was *appet*". As a matter of fact, some participants, when performing the free memory recall tasks and the picture naming tasks, instead of writing or pronounce the correct new pseudo-word, provided in few cases its base word instead. This is in line with the strategy we propose here, with the participants referring to the already existing integrated word when trying to recall the newly learned similar pseudo-word.

4.2 Neural findings

We hypothesized that the retrieval of words of the Recent condition would rely more on a hippocampal memory representation as stated by the CLS account, and words of the Remote condition with a memory representation in the pMTG as the previous literature suggested. Contrary to our hypothesis we did not find such a pattern. When comparing the Recent and the Remote conditions we found the hippocampus active in both Remote and Recent condition. In the Remote condition, rather than pMTG we found activation of the IFG which, in the MUC (Hagoort, 2013), is considered to be an area responsible for linguistic unification processes. Among the others, IFG is supposed to be a key node in the semantic unification network, unifying semantic information from different modalities. In addition, the pMTG has been found active in the remote condition when we lowered the cluster threshold.

4.2.1 Correlational Analyses

We tested to see if the activation pattern in the contrast Remote minus Recent changed as a function of proficiency of L2, dictionary size in L2, and length of stay in a L2 country, but no significance correlation has been found in the relevant contrasts. The activation of hippocampus and pMTG did not change as function of these individual differences, hence, it seems that these covariates do not have influence on the consolidation process of new words in L2.

4.2.2 Imaging

We wanted to investigate which area is responsible to connect the word-form to its conceptual meaning, which is thought to be the pMTG (Hagoort, 2013; Hickok & Poeppel, 2004), and the role of sleep and consolidation in the shift from an hippocampal episodic memory to an integrated cortical representation (Davis & Gaskell, 2009). The imaging results, both at the whole brain analysis and ROIs analysis, did not reveal the expected pattern of activation. Contrary to our hypothesis, hippocampus have been found active in both Remote and Recent condition, while in the remote condition IFG has been found active, with an under threshold activation of the pMTG. In spite of that, the behavioral effects of integration arouse.

Hence, in contrast with our hypothesis, sleep but not consolidation appears to be beneficial for the occurrence of lexicalization and semanticization processes. It seems that the behavioural effects of integration arouse in spite of a missed occurrence of a complete integration process. It could be the case that the consolidation process is slower than the integration process, but still strong enough to lead to behavioural differences as suggested by the hippocampal activation and the under-threshold activation of the pMTG in the Remote condition. It could be the case that the consolidation process in L2 takes longer, given a smaller vocabulary than L2, which the literature (James et al., 2017; van Goch et al., 2017) as our data, show to be an important aspect for word learning.

IFG activation is in line with previous literature which shows the IFG as an on-line construction of a new and unified representation of the input streams. (Willems, Özyürek, & Hagoort, 2009). Since in the task participant had to recognize the object in a different picture and associate it to the correct word-form, IFG activation seems to be related to the task design.

In addition, it could be that the behavioral measurements of integration are more sensitive than the fMRI imaging of consolidation. Hence, we are able to detect integration effects but not the underpinning brain consolidation process responsible for that. Alternatively, it could be the case that there is no one-to-one correspondence between consolidation and integration.

In addition, the literature on lexicalization (Kapnoula & McMurray, 2016; Kapnoula et al., 2015; Lindsay & Gaskell, 2013.; Szmalec et al., 2012) shown that to provide the word-form can have some effect in the processes discussed in this research. In our task performed in the scanner, we provided the correct word-form among the possible options. We did not choose a picture naming task in the scanner, since the motion caused by the vocalization would spoil the imaging data, making their analysis more difficult. Considering that we provided the word-form in the fMRI task, it is possible that the provision of the word-form triggered a hippocampal episodic memory of learning the word-form. The memory of learning the word is episodic and still based on an hippocampal representation. This could explain why we found hippocampal activation not only in the Recent condition but also in the Remote condition. This explanation is also in line with the multiple trace theory which posits that the hippocampus retains a permanent role in memory storage and retrieval as long as memories exist (Moscovitch et al., 2005, 2006; Nadel & Moscovitch, 1997) and furthermore that the retrieval of preexisting knowledge could occur without the contribution of the medial temporal lobe memory system.

4.3 Further researches

Several questions arise from this research, about the role of proficiency, the provision of word-forms, and consolidation in L2.

In order to make our results about the role of vocabulary size more robust, we could repeat this experiment using the training procedure and the free memory recall task. This time, however, we would use the Peabody picture vocabulary test (PPVT) in order to control the vocabulary size (Miller & Lee, 1993). PPVT is a more reliable measurement of the dictionary size than the measure we used in this research.

If the proficiency plays a role in word learning with a critical time point after which it is not anymore involved, we could design an experiment to investigate this issue. We could take a group with an already high proficiency in a determined L2 and see their performance in a free memory recall task, after learning some L2 words. Then we repeat the procedure with an unknown L3. If the number of words remembered in the two conditions is different, with a higher number of L2 remembered words, it would be possible that proficiency plays a role. Furthermore we could investigate which aspects of proficiency are important in the crossing of the critical point. Probably, a key factor of proficiency is the phoneme categorization in a foreign language. Hence, we could control for the phoneme categorization in L1, L2, and L3 and select specific words of L3 which have contrast not present in L1 or L2. If the words of L3 which require an unknown phoneme categorization are remembered less, we could argue that proficiency and specifically, phoneme categorization, plays a role in word learning. On the other hand, if we do not found any significant difference between the numbers of L2 and L3 remembered words, we can strongly argue against a role of proficiency in L2 word learning.

About the role of the provision of base-word in the detection of competition effect and lexicalization we could use both lexical competition task and the visual word paradigm. Then, we could control for any difference between the results in the two tasks. After teaching of some new words, we could test participants immediately after training, after 12 hours without sleep, and 24 hours later with a sleep consolidation period. If immediately after training the visual word paradigm shows some competition effects, but not the lexical competition task, we could argue that the visual word paradigm fails to detect the competition which arise from lexicalization but instead detect a competition which arise from an episodic memory trace. Furthermore, the results would be also informative of the time scale of integration processes.

To control if consolidation in L1 and L2 follows a different time scale, we could design an experiment similar to the one presented here, but this time with a set of words from L1 and another one from L2. We could then investigate with the fMRI scanner, the same day of training, the next day, and after a week, if the pattern of activation differs between L1 and L2. This would provide a clear information about any difference between L1 and L2 in the consolidation process, and also more general information of the time scale of the consolidation process.

5 Conclusions

Word learning appears to occur in a similar manner in L1 and L2. The lexicalization process is perfectly in line with the previous literature in L1 with competition only after sleep. Also the semanticization process shows a trend in line with the previous literature in L1, with the priming effect mainly driven by the words which undergone two nights of sleep. Given our results no strong conclusion can be drawn on semanticization but nevertheless the data seems to point towards a similar semanticization process in L1 and L2. In addition, word learning in L2 appears to be not influenced by the proficiency in L2 as by the length of stay in an L2 country. Together, the results suggest that word-learning is a process which does not change between L1 and L2. The only individual difference which seems to be beneficial to word learning appears to be the vocabulary size. Hence, knowing more word in a language should help the memorization of new words. The consolidation process seems to not have a one-to-one correspondence with lexicalization and semanticization and a higher dependency on the hippocampus for both the Remote and Recent condition.

Acknowledgements

Thanks to James, who believed in my research and gave me this amazing opportunity to test my ideas. Thanks to Atsuko for being extremely patient with me, I know that my communication is not always ideal. Thanks to my family which always supported me. Thanks to my mum who made me smart and to my father who made me beautiful. Thanks to Tobia for his amazing contribution. We will "quench" another time together, spinning on the chairs around Donders. Thanks to all the Italians that I tested and allowed me to graduate. Without you, I would never have been able to "discover how Italians learn English words". But as I said: "Italian community is strong"; and you proved it! Thanks to Betty & Mora, without their real espresso I would never have been productive, and neither able to convince people to participate in my study. Thanks to all the friends who saved me. Thanks to the "fellowship", Jordi, Adrian, and Jacob. Without you guys, I would be mad now (I mean, even madder). Thanks to the Sceriffo Giuseppe, and all the "gang of Bang", Michele, Thomas, Flavia, Fabio, Giorgia, Lorenzo, Andrea. We will play again. Thanks to Laura for being "my mum abroad". Thanks to all the people from Vossendel, we learned what does it mean to bike regardless of the weather. Thanks to the football players of the Sunday, one day I will score. Thanks to Marco for being always there when needed. Thanks to Pepla for his always interesting discussions, we will destroy structuralism sooner or later (probably later but we will, actually shall). Thanks to Nader for being crazy. Thanks to all the ladies who donated me a smile. Thanks to GNU Sound and Croce d'Oro, the associations which formed me creating the person that I am now. Thanks to the only real traditional Dutch food: the herring, which save me from several hangovers (no, bitterballen does not count as traditional, everybody can do bitterballen). Thanks to the Netherlands, it has been though but we accepted each other (more or less).

"Step by step"

~~Confucius~~

Giacomo

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, *38*(4), 419-439.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 802.

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*, 116-130.

Bardovi-Harlig, K., & Bastos, M. T. (2011). Proficiency, length of stay, and intensity of interaction and the acquisition of conventional expressions in L2 pragmatics. *Intercultural Pragmatics, 8*(3), 347-384.

Barron, A. (2003). Acquisition in interlanguage pragmatics: Learning how to do things with words in a study abroad context (Vol. 108). *John Benjamins Publishing.*

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in cognitive sciences*, *15*(11), 527-536.

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.

Breitenstein, C., Zwitserlood, P., de Vries, M. H., Feldhues, C., Knecht, S., & Dobel, C. (2007). Five days versus a lifetime: Intense associative vocabulary training generates lexically integrated words. *Restorative neurology and neuroscience*, *25*(5, 6), 493-500.

Brysbaert, M. & New, B. (2009) Moving beyond Kucera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. Behavior Research Methods, 41 (4), 977-990.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature*, *349*(6312), 788.

Casey, S. J., Solomons, L. C., Steier, J., Kabra, N., Burnside, A., Pengo, M. F., ... & Kopelman, M. D. (2016). Slow wave and REM sleep deprivation effects on explicit and implicit memory during sleep. *Neuropsychology, 30*(8), 931.

Charge, N., & Taylor, L. B. (1997). Recent developments in IELTS. *ELT Journal*, *51*(4), 374-380.

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.

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 551.

Cohen, A. D., & Aphek, E. (1981). Easifying second language learning 1. *Studies in second language acquisition*, *3*(2), 221-236.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive psychology*, *42*(4), 317-367.

Davis, M. H., Di Betta, A. M., Macdonald, M. J., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of cognitive neuroscience*, *21*(4), 803-820.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35-39.13 Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010

Dumay, N., Frauenfelder, U. H., & Content, A. (2000). Acoustic-phonetic cues and lexical competition in segmentation of continuous speech.

Fan, M. Y. (2003). Frequency of use, perceived usefulness, and actual usefulness of second language vocabulary strategies: A study of Hong Kong learners. *The Modern Language Journal*, *87*(2), 222-241.

Ferrand, L., & New, B. (2003). Semantic and associative priming in the mental lexicon. *Mental lexicon: Some words to talk about words*, 25-43.

Ferré, P., Sánchez-Casas, R., & Guasch, M. (2006). Can a Horse Be a Donkey? Semantic and Form Interference Effects in Translation Recognition in Early and Late Proficient and Nonproficient Spanish-Catalan Bilinguals. *Language Learning*, *56*(4), 571-608.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1187.

Frauenfelder, U. H., & Peeters, G. (1990). Lexical segmentation in TRACE: An exercise in simulation.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in second language acquisition*, *26*(2), 275-301.

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

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of memory and language*, *28*(5), 501-518.

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

Guasch, M., Sánchez-Casas, R., Ferré, P., & García-Albea, J. E. (2008). Translation performance of beginning, intermediate and proficient Spanish-Catalan bilinguals: effects of form and semantic relations. The Mental Lexicon, 3(3), 289-308.

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.

Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and cognitive processes*, *8*(4), 379-411.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and cognitive processes*, *5*(4), 281-312.

Kapnoula, E. C., & McMurray, B. (2016). Newly learned word forms are abstract and integrated immediately after acquisition. *Psychonomic bulletin & review*, *23*(2), 491-499.

Kapnoula, E. C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, *134*, 85-99.

Karni, A., Tanne, D., Rubenstein, B., Askenasy, J., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science*, *265*(5172), 679–682. https://doi.org/10.1126/science.8036518

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods 42*(3), 627-633.

Krieger-Redwood, K., Teige, C., Davey, J., Hymers, M., & Jefferies, E. (2015). Conceptual control across modalities: graded specialisation for pictures and words in inferior frontal and posterior temporal cortex. *Neuropsychologia*, *76*, 92-107.

Kroll, J. F., & Sholl, A. (1992). Lexical and conceptual memory in fluent and nonfluent bilinguals. In Advances in psychology (Vol. 83, pp. 191-204). North-Holland.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. Journal of memory and language, 33(2), 149-174.

Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection?. *Journal of Memory and language*, *41*(3), 365-397.)

Lindsay, S., & Gaskell, M. G. (2010). A complementary systems account of word learning in L1 and L2. *Language Learning*, *60*(s2), 45-63.

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 608.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological science*, *20*(12), 1507-1515.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19*(1), 1.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*(6), 485-495.

Marian, V., Bartolotti, J., Chabal, S., Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE 7(8): e43230*.doi:10.1371/journal.pone.0043230

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71-102.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human perception and performance*, *15*(3), 576.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Current opinion in neurobiology*, *11*(2), 194-201.

Mattys, S. L., & Clark, J. H. (2002). Lexical activity in speech processing: Evidence from pause detection. *Journal of Memory and Language*, *47*(3), 343-359.

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. https://doi.org/10.1037//0033-295X.102.3.419

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.

McKoon, G., & Ratcliff, R. (1979). Priming in episodic and semantic memory. *Journal of verbal learning and verbal behavior, 18*(4), 463-480.

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(3), 621.

Miller, L. T., & Lee, C. J. (1993). Construct validation of the Peabody Picture Vocabulary Test—Revised: A structural equation model of the acquisition order of words. *Psychological Assessment, 5*(4), 438.

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.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1209.

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.

Pajak, B., Creel, S. C., & Levy, R. (2016). Difficulty in learning similar-sounding words: A developmental stage or a general property of learning?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(9), 1377.

Pavlenko, A. (Ed.). (2009). *The bilingual mental lexicon: Interdisciplinary approaches* (Vol. 70). Multilingual Matters.

Perea, M., Dunabeitia, J. A., & Carreiras, M. (2008). Masked associative/semantic priming effects across languages with highly proficient bilinguals. *Journal of Memory and Language*, *58*(4), 916-930.)

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

Ripollés, P., Biel, D., Peñaloza, C., Kaufmann, J., Marco-Pallarés, J., Noesselt, T., & Rodríguez-Fornells, A. (2017). Strength of Temporal White Matter Pathways Predicts Semantic Learning. *Journal of Neuroscience, 37*(46), 11101-11113.

Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annu. Rev. Neurosci.*, *27*, 279-306.

Stickgold, R. (1998). Sleep: off-line memory reprocessing. *Trends in Cognitive Sciences*, *2*(12), 484–492. https://doi.org/10.1016/S1364-6613(98)01258-3

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1272–1278. https://doi.org/10.1038/nature04286

Szmalec, A., Page, M. P., & Duyck, W. (2012). The development of long-term lexical representations through Hebb repetition learning. *Journal of Memory and Language*, *67*(3), 342-354.

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in second language acquisition, 28*(3), 387-422.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition, 2*(1), 45-58.

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., 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*, 265-278.21 Tamminen & Gaskell, 2013;

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., 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.

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. *Journal of Neuroscience*, *30*(43), 14356-14360.

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.

Traxler, M. J. (2011). *Introduction to psycholinguistics: Understanding language science*. John Wiley & Sons.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and cognition*, *4*(2), 105-122.

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.

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.

Van Goch, M. M., Verhoeven, L., & McQueen, J. M. (2017). Trainability in lexical specificity mediates between short-term memory and both vocabulary and rhyme awareness. *Learning and Individual Differences*, *57*, 163-169.

Ven, F., Takashima, A., Segers, E., & Verhoeven, L. (2017). Semantic priming in Dutch children: Word meaning integration and study modality effects. *Language Learning*, *67*(3), 546-568.

Vroomen, J., & De Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human perception and performance*, *21*(1), 98.

Walker, M. P., & Stickgold, R. (2004). Sleep-Dependent Learning and Memory Consolidation. *Neuron*, *44*(1), 121–133. https://doi.org/10.1016/j.neuron.2004.08.031

Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with Sleep Makes Perfect. *Neuron*, *35*(1), 205–211. https://doi.org/10.1016/S0896-6273(02)00746-8

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1-25. &&&& Costa, A., Miozzo, M., & Caramazza, A. (1999).

Willems, R. M., Özyürek, A., & Hagoort, P. (2006). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*, *17*(10), 2322-2333.

Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience*, *20*(7), 1235-1249.

Willems, R. M., Özyürek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage*, *47*(4), 1992-2004.