Master's Thesis
in Language and Communication
(Research)

Effects of consolidation in foreign language attrition

Ekaterina Slesareva

July 22nd, 2019

Supervisors:
Dr. K.M. Lemhöfer
Prof. Dr. J.M. McQueen
A. Mickan (M.Sc.)

DONDERS INSTITUTE

Abstract

Effects of consolidation in foreign language attrition

There is preliminary evidence suggesting that learning of a new foreign language (L3) leads to interference in a previously learned foreign language (L2), as revealed by higher naming latencies for L2 words (Mickan et al., in preparation). The present study explored whether offline, overnight consolidation makes a difference in the extent to which the newly learned words in a new foreign language interfere with their translation equivalents in already known foreign languages. Participants, who had no knowledge of Spanish, first underwent an English vocabulary test, which resulted in a list of well-known English words. Half of the words on that list were learned in Spanish. In a final test, participants had to name all the English words from the list again. Importantly, one group of participants learned new words in Spanish and had the final English test on the same day. The other group learned new words in Spanish in one day and had their knowledge of English words tested 24 hours afterwards, which gave time to the newly learned words to be consolidated overnight. In regard to the interference effects, it was found that learning of an L3 led to interference in an L2, which was seen in both accuracy and naming latencies, meaning that learning of an L3 may actually result in forgetting of an L2. In regard to consolidation, the results tentatively suggest that it leads to a stronger interference effect in accuracy, but not in naming latencies.

Table of Contents

Introduction

I believe that many people around the world have experienced the following: a person learns a foreign language at school, after which they begin learning another foreign language because it might be needed for their job or because they have moved to another country. At that point people often feel that the already known foreign language is being replaced by a new foreign language. But is it actually the case? Does learning a new foreign language make you forget previously learned foreign languages? In this thesis, I will shed light on this question and, more generally, attempt to further our understanding of how multiple languages are handled within one's brain and how these languages influence one another and interact.

To start with, there is preliminary evidence suggesting that learning a new foreign language indeed impairs the retrieval of previously learned and well established foreign languages. This evidence comes from research by Mickan, McQueen & Lemhöfer (in preparation). Their participants, who had no knowledge of Spanish, were first tested in English (L2) via a picture naming task. Based on the participants' correct answers in this English test, a list of well-known English words was created. Half of the words on that list were then learned in Spanish (L3), and the other half were not. In a final English test, participants had to name all the English words from the list again. To investigate whether Spanish learning had any impact on the participants' English, the researchers compared accuracy as well as naming latencies for words previously learned in Spanish to words that were not. The results showed significantly higher naming latencies in English for words that were also known in Spanish. However, the researchers found no effect for English naming accuracy, meaning that participants did not make more errors in words that were learned in Spanish.

Mickan et al. (in preparation) explained their results in terms of the interference theory. The interference theory originates from memory research, where it has been used to explain one of the most basic problems that the field is confronted with: forgetting. According to the interference theory, forgetting of a certain memory happens due to competition, and thus interference, from other related memories, i.e. memories that overlap in content and, thus, respond to the same retrieval cues (Keppel, 1968; Postman & Underwood, 1973). Namely, in the process of retrieval, connected memories are co-activated, they compete for selection and, as a result, hinder each other's retrieval. Moreover, repeated retrieval of a certain memory is linked to strengthening of that memory, but it comes at cost of simultaneous breakdown of access to other connected, but non-retrieved memories. It is argued that this access breakdown occurs due to the active inhibition of non-retrieved memories needed for the retrieval of the target memory. The impairment can even become long-lasting, which will lead to a memory failure in the shape of forgetting. This phenomenon is referred to as retrieval-induced forgetting (RIF) (Anderson, Bjork, & Bjork, 1994; Anderson & Spellman, 1995).

The interference theory and RIF are also believed to be applicable to language systems in general and language attrition in particular. Previous research has reliably demonstrated that languages that a person speaks are closely related within the brain (e.g. Dijkstra, Grainger & van Heuven, 1999; Kaushanskaya & Marian, 2007; Bergmann, Sprenger, & Schmid, 2015). On the word level, translation equivalents in different languages are connected to one another via the shared concept, which leads to competition during speech production (e.g. Dijkstra, 2005). For instance, if a bilingual speaker is asked to name a picture (e.g. a picture depicting a "die") in a certain language (e.g. Spanish), the lexical concept, i.e. the meaning of "die", results in activation of two lemmas with two different language tags (e.g. "die" with an English tag and "dado" with a Spanish tag) and in their competition at the retrieval. According to the interference theory, language attrition is directly linked to the increasing impact of, or interference from, the competing languages (for an overview see Bardovi-Harlig & Stringer, 2010). Thus, as mentioned above, Mickan et al. (in preparation) used the interference theory to explain the findings of their study. Specifically, Spanish learning in their study was looked at as interference. Higher naming latencies for the English items that participants also learned in Spanish were claimed to be the result of the interference from Spanish. That is, Spanish competitor words, that were learned during the Spanish learning phase, interfered with the subsequent retrieval of the target English words, and this interference effect was manifested in the fact that participants took longer to name English items that they also knew Spanish labels for. Moreover, this finding can also be explained as an instance of RIF. Since Spanish words were retrieved multiple times during the Spanish learning phase, the retrieval of English words that are connected semantically was ultimately impaired at the final test, and participants needed more time to produce the word.

There is more evidence that links between-language competition and interference to forgetting. For instance, it has been documented in studies on the influence of L2 on L1. In the study by Levy, McVeigh, Marful and Anderson (2007) participants were asked to repeatedly name pictures in either English, their L1, or Spanish, their L2. In a final test, participants had to name all the pictures in English. It was observed that while repeated naming of the picture in English increased the recall of the corresponding word in the final test, the repeated naming of the picture in Spanish, i.e. more than one time, significantly impaired the subsequent recall of its translation equivalent in English in the final test.

Furthermore, in the experiment by Isurin & McDonald (2001) participants studied a number of words (concrete objects and actions) in Russian or Hebrew (referred to as L1 by researchers, although the participants' actual L1 was English). After that, participants learned another list of words in a different language (referred to as L2). Words in the second list were either translation equivalents of words in the first list or words that did not overlap in meaning with the first list. Researchers found more interference, expressed both in lower accuracy and slower reaction times in the L1 retest, for translation equivalents compared to novel words as well as for increased number of exposures to the second list.

Finally, the connection between competition, interference and forgetting has also been shown in studies of L1 and L2 influence on L3 learning. Mickan, McQueen, & Lemhöfer (submitted) taught participants a number of words in Spanish (L3). Translation equivalents of half of the learned Spanish words were then used for tasks in Dutch (participants' L1) or English (participants' L2). Finally, participants were again tested on all the Spanish words. Researchers compared naming latencies and accuracy in the last Spanish test for items interfered with Dutch or English to non-interfered ones, and they found significant interference effects, manifested both in naming latencies and accuracy.

All in all, findings discussed above indicate that languages that a person knows influence one another bidirectionally, which results in between-language competition, interference and even forgetting in the long term. Most importantly for my thesis, there is preliminary evidence for the fact that learning a new foreign language might interfere with previously learned foreign languages, but only to some extent, i.e. only in naming latencies as shown by Mickan et al. (in preparation). Thus, Mickan et al. (in preparation) did not find that people actually "forgot" English words in laymen interpretation of the term. However, in this study, at the moment of the final English test, novel Spanish words have just been learned, since all the experimental manipulations took place within one day. Hence, it might be possible that novel words in L3 were too weak to interfere considerably. By comparison, Mickan et al. (submitted) have shown clear forgetting effects in naming latencies as well as in accuracy, since in their study interference came from well-known and established languages (L1 and L2). Thus, one might speculate that if in the study by Mickan et al. (in preparation) Spanish words were given more time to stabilise in the participants' minds, the researchers would also observe that these new L3 words might turn out to be stronger interferers.

In fact, research on novel word learning has clearly shown that it takes time for novel words to be integrated into the mental lexicon and to compete with other words in the system. Namely, a model, called Complementary Learning Systems (CLS), has been proposed by Davis and Gaskell (2009) to account for the process of novel word learning. It builds on the theory of systems-level consolidation coming from memory research (Marr, 1970; McClelland, McNaughton, & O'Reilly, 1995). According to the CLS model, word learning consists of two stages: hippocampal learning that is followed by gradual overnight word consolidation. Namely, new memories, i.e. new words, are first encoded as episodic memory traces by the hippocampal system that subserves fast learning. The episodic representations make immediate recognition and production of the newly learned words in explicit memory tasks possible. However, these representations exist in a relative isolation from the existing lexicon which is stored in neocortex. Reactivation of hippocampal memory traces solidifies connections between neocortical memory components. It is known that hippocampal memory traces are reactivated during post-learning rest, especially sleep (Frankland & Bontempi, 2005). As connections in neocortex grow stronger, the hippocampal activity reduces continuously until the memory trace is fully independent of hippocampus. This slow integration is also known as consolidation and is thought to be crucial to avoid catastrophic interference – the breakdown of existing connections due to new information being presented (McClosky & Cohen, 1989).

The support for the CLS model was found in both behavioural and neuroscientific experiments on novel word learning. For instance, Dumay and Gaskell (2007) investigated the role of sleep in consolidation of novel words. In their experiment, participants learned a number of nonsense competitor words (e.g. *shadowks*) either in the morning or in the evening. They were asked to come back 12 hours later to have their lexical activity (via pause detection) and explicit memory (via free recall, 2-AFC recognition) measured. Thus, the group who learned the novel words in the morning returned to the lab on the same day, while the group who learned the words in the evening returned to the lab on the next day. People in the morning group, who did not sleep, did not exhibit competitions effects neither immediately after the learning sessions nor after 12-hour rest interval. However, the evening group, whose 12-hour rest interval included sleep, showed competition effects at the time of the second testing, but not right after the learning session. Consequently, it can be concluded that novel words do not enter the lexical competition neither right after learning nor after a rest period which does not include sleep, but they do after a night's sleep. Thus, the major implication of this study was that consolidation of new words is closely related to sleep.

As far as the neuroscientific evidence is concerned, Takashima, Bakker, Van Hell, Janzen, and McQueen (2014) carried out an fMRI study in which participants had to learn a set of words and were tested on these immediately after the learning session and 24 hours after. When the delayed test was compared to the immediate test, the increased activity in bilateral pMTG, which is a region involved in lexical storage, as well as stronger connectivity between pMTG and the auditory cortex were observed. In another fMRI study by Davis, Di Betta, Macdonald, and Gaskell (2009) participants were trained on two sets consisting of nonsense and real words on two consecutive days (one set was trained on day 1, another set was trained on day 2). After the training session on the second day, participants performed a pause detection task while their brain activity was measured. Importantly, items that participants were not trained on at all were also included in the task. Results showed an increased hippocampal activity for untrained words and decreased hippocampal activity for unconsolidated words, i.e. those that were trained on day 2. Cortical activation for both untrained and unconsolidated words was of the similar degree. In contrast, decreased neocortical activity was seen for consolidated new words, i.e. those that were learned on day 1, meaning that these words started acting similarly to other words in the lexicon.

To sum up, the aforementioned studies have determined that newly learned words do not necessarily become a part of the language network right away: a period of offline consolidation might be needed for that (but see Borovksy, Elman, & Kutas, 2012; Lindsay & Gaskell, 2013; Coutanche & Thompson-Schill, 2014, for evidence of rapid lexicalisation effects). It has been found that a night's sleep aids the consolidation of novel words. Applied to the study by Mickan et al. (in preparation), it means that the Spanish words that participants learned were not fully consolidated, since the consolidation process takes time. So, it is unclear whether consolidation and lexicalisation of the Spanish words can result in even stronger competition between English (L2) and Spanish (L3) words. Thus, the goal of the current study is to explore the following: does extra, offline consolidation (including a night of sleep) make a difference in the extent to which the newly learned words in a new foreign language interfere with their translation equivalents in other, previously learned foreign languages?

The experimental procedure is similar to Mickan et al. (in preparation): first, participants (monolingually raised Dutch native speakers with good knowledge of English, but no knowledge of Spanish) undergo an English vocabulary test. Based on the participant's correct answers, a list of words is created. Half of the list is subsequently learned in Spanish. In the final test, the participants are asked to name all the words from the list in English. The interfered words, i.e. those taught in Spanish, are compared to the non-interfered words with regard to naming accuracy and naming latency in English. Crucially, there are two groups of participants. One group learns new words in Spanish and has the final English test on the same day, similarly to Mickan et al. (in preparation). The other group learns new words in Spanish in one day and has their knowledge of English words tested 24 hours afterwards (which will give time to the newly learned Spanish words to be consolidated overnight). The two groups are then compared in order to answer the research question. The present study introduces one more important contrast in regard to the experiment by Mickan et al. (in preparation) in addition to the presence of two consolidation groups. In the previous experiment the English pre-test took place the same day as the Spanish learning session, however, in this experiment it happens one day beforehand. Re-activation of memories is thought to make memory entries vulnerable (Spear, 1973; Nader & Hardt, 2009). Consequently, having the English pre-testing occur immediately before the Spanish learning stage may have resulted in greater impairment in the English retrieval in the study by Mickan et al. (in preparation).

Based on the previous studies by Mickan et al. (submitted; in preparation), we hypothesise to observe higher naming latencies for the interfered words in both groups. Moreover, greater interference is expected to be found in the consolidation group, since Spanish words are supposed to enter the lexical competition after the night's sleep, causing reaction times to become even slower and possibly affecting the accuracy.

Methods

Participants

Originally, 55 people participated in the study. However, the data of one participant had to be excluded. That participant did not learn enough Spanish words during the Spanish learning session (see the section Exclusion criteria for more details). Thus, 54 participants (17 males and 37 females), aged 18-29 years (*M* = 21.91, *SD* = 2.76), participated in the experiment.

Participants were recruited via SONA, the Radboud research participation system. The eligibility of every participant was verified via an online language background questionnaire that had to filled in before the participant could sign up for lab sessions. The participants' only mother tongue had to be Dutch, their age had to be between 18 and 35 years, and they had to have very little (3 participants) to no knowledge of the Spanish language (51 participants).

All participants who had little knowledge of Spanish indicated that they almost never used Spanish in their everyday lives. The participants were 19, 20 and 23 years old, and they learned some Spanish at the age of 9, 19 and 16, respectively. Self-ratings of these participants for their Spanish language skills (speaking, listening, writing, and reading) on a scale from 1 (very poor) to 7 (very good) were the following: the first participant rated all skills as 1; the second participant rated listening and reading skills as 2, speaking and writing as 1; the third participant rated all skills as 2.

Moreover, the participants' English proficiency had to be intermediate/advanced. We controlled for it by limiting the participation of people who were enrolled in English-taught study programs (all but one participant were studying in Dutch) and people who spent to an English-speaking country for more than 3 months (all but one participant have not spent more than 3 months in an English-speaking country).

In the experiment, participants were randomly assigned to one of two groups: one group of the participants (no consolidation group) learned new words in Spanish and underwent Spanish and English post-tests on the same day. The other group (consolidation group) learned new words in Spanish in one day and had their knowledge of interfered English items tested 24 hours afterwards (which gave time to the newly learned Spanish words to be consolidated overnight).

The participants in the two groups were compared by means of independent t-tests on a number of dimensions that could influence the degree of interference in their L2 (see Table 1). English vocabulary size was determined with the help of the LexTALE vocabulary test (Lemhöfer & Broersma, 2012). All variables except LexTALE scores were collected through a language background questionnaire.

Table 1. *Ranges, mean scores and standard deviations of participant characteristics in the two groups.*

	Consolidation group *n* = 27			No Consolidation group *n* = 27			Test statistics
	Range	*Mean*	*SD*	*Range*	*Mean*	*SD*	
Age	18-29	21.81	3.06	18-27	22.00	2.47	$t(49.72) = -0.24$, $p = 0.81$
Age of acquisition (English)	6-14	10.37	1.71	7-14	11.15	1.59	$t(51.69) = -1.73$, $p = 0.09$
Years of learning English	3-22	9.93	4.80	4-14	9.56	2.55	$t(39.55) = 0.35$, $p = 0.73$
Self-rated English proficiency: Speaking*	3-7	5.48	0.94	2-7	4.93	1.04	$t(51.48) = 2.07$, $p = 0.04$
Self-rated English proficiency: Listening*	4-7	6.22	0.80	2-7	5.59	1.12	$t(47.10) = 2.38$, $p = 0.02$
Self-rated English proficiency: Reading*	3-7	6.04	0.94	3-7	5.74	0.94	$t(51.99) = 1.16$, $p = 0.25$
Self-rated English proficiency: Writing*	3-7	5.37	0.93	2-7	4.70	1.07	$t(50.98) = 2.45$, $p = 0.02$
Vocabulary size: LexTALE score**	52-95	77.15	11.47	51-92	73.52	10.42	$t(51.52) = 1.22$, $p = 0.23$

Note. *SD* = Standard Deviation. Variables marked with one asterisk were self-rated on a 1–7 (1 = *very poor*, 7 = *very good*) scale. Variables marked with two asterisks indicate a percentage. Ranges show the absolute min and max values per group and condition.

The study was approved by the Ethics committee of the Faculty of Social Sciences of Radboud University. Each participant signed an informed consent form and a screening form before the start of the experiment. They received a reimbursement of 10 euros or 1 participation point per hour (max. 25 euros or 2.5 participation points for three sessions). The payment depended on how much time the participant spent in the lab.

Materials

A total number of 140 everyday words were used in the study (the list can be found in Appendix). The characteristics of the chosen words were the following.

- All words were non-cognates in Spanish, English and Dutch. Importantly, we avoided identical as well as non-identical cognates.

- The words were not compounds in Spanish or English.
- Word length (in syllables) ranged from 1 to 5 in Spanish (*M* = 2.56, *SD* = 0.69) and from 1 to 4 in English (*M* = 1.42, *SD* = 0.60).
- The frequency of occurrence ranged from 0 to 818 per million in Dutch (*M* = 29.02, *SD* = 76.69) and from 0 to 354 per million in English (*M* = 31.3, *SD* = 45.26) according to the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995).

Coloured photographs corresponding to these everyday words (400 pixels wide x 400 pixels high) were used as visual stimuli for all tasks. All of them had been found on the internet (Google Images and Flickr). The pictures were shown to participants on a white background on a computer screen (Benq XL2420Z, 24 inch). The picture was presented for the whole duration of the trial. Moreover, participants listened to recordings of the words during the Spanish learning phase via headphones (Sony MDR-7506). The audio files had been recorded by a female native (peninsular) Spanish speaker.

The participants' responses were recorded with the Shure SM-57 microphone and the Behringer X-Air XR18 digital mixer. The recording for each trial started at the onset of presentation of a visual stimulus and continued for 30 seconds.

Procedure

As outlined in the introduction, the experiment consisted of three parts that took place on either two or three consecutive days depending on the group the participant was in. The first part was dedicated to testing the participant's knowledge of English words; the participant learned Spanish words in the second part and had his/her knowledge of these words tested; in the third part, the participant's English was tested once again.

The study occured in the same sound-proofed lab at the Donders Centre for Cognition for all participants. The lab consisted of two rooms: controller room and participant room. The controller room, where the experimenter stayed for the whole duration of the experiment, was adjacent to the participant room. The door between these two rooms was kept open during the experiment to allow for coding of the participants' answers. The participants were seated at the arm's length distance (50 cm) from the computer screen. All participants were tested individually. The experiment was run with the use of the software package Presentation (Version 17.0, Neurobehavioral System Inc, Berkeley, U.S.).

Day 1: English Pre-test

When the participant came to the lab on the first day, they were told that even though the experiment was about learning Spanish, we would start with testing their English. In the English pre-test, the participant saw 140 pictures on the screen shown one by one, which they needed to name in English. In case they did not know the word, they were asked to say "I don't know", and then the next picture was shown. They were asked to avoid using articles or full sentences while naming. The experimenter coded the answers as correct or incorrect. If a synonym was used, the answer was coded as incorrect. After this task, the participants were told whether they would need to come back to the lab for one more day or two more days (no consolidation/consolidation group), and then they could go home.

Item Selection

After the pre-test, a script written in MATLAB (Version 9.5; The MathWorks, Inc., 2018) was run in order to create lists of words that would be used for all subsequent tasks. Initially, we chose the "standard" list consisting of 46 words, which were selected for the experiment in case the participant could name them in the pre-test. Each item in the list was assigned to one of the two conditions: an interference condition (23 items), meaning that the item would later be learned in Spanish, and a non-interference condition (23 items), meaning that the item would not be learned in Spanish. We counterbalanced the assignment of these items to conditions between participants. If an item from the standard list was not known to the participant, the script looked for a substitution among the items that the participant knew and that matched item characteristics as closely as possible. Items across and within both conditions were matched on:

- word length (in syllables) in Spanish and English;
- the frequency of occurrence in Dutch and English according to the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995);
- semantic similarity (as described by Mandera, Keuleers, & Brysbaert, 2017);
- phonological similarity in Spanish (Levensthein distance ≥ 2 between any two words of the entire set).

That way, the created lists of words were participant-specific. More details on item characteristics of the ultimate lists that were used in the experiment can be found in Table 2.

Table 2. *Characteristics of items used in the experiment.*

	Consolidation group						**No Consolidation group**					
	Interference set			No interference set			Interference set			No interference set		
	Range	*Mean*	*SD*	*Range*	*Mean*	*SD*	*Range*	*Mean*	*SD*	*Range*	*Mean*	*SD*
English word length (in syllables)	1-3	1.34	0.49	1-3	1.39	0.49	1-4	1.38	0.51	1-3	1.37	0.49
English Celex log frequency	0-2.55	1.27	0.46	0-2.55	1.28	0.48	0-2.55	1.29	0.50	0-2.55	1.30	0.48
English Celex per million frequency	0-354	31.18	37.85	0-354	33.83	43.36	0-354	35.51	46.17	0-354	35.61	46.42
Dutch Celex log frequency	0-2.32	1.01	0.57	0-2.91	1.00	0.60	0-2.91	1.04	0.60	0-2.31	1.05	0.58
Dutch Celex per million frequency	0-188	14.52	21.68	0-188	15.09	22.90	0-188	20.05	34.40	0-188	18.09	31.45

Note. *SD* = Standard Deviation. Items sets differed across participants, as described in the Item selection section. Ranges show the absolute min and max values per group and condition.

Day 2: Spanish Learning Phase

The comprehension and production tasks for learning Spanish that will be described below only involved 23 words from the interference condition. The order of presentation of the words was kept constant within each task, but differed between the tasks.

Spanish familiarisation task. In the Spanish familiarisation task, participants were presented with 23 pictures shown one by one. Participants heard the Spanish word corresponding to the picture via the headphones and also saw the spelling of the word (font Verdana, size 20) on the screen below the picture. They were also asked to repeat the word out loud after they heard it. Participants could move to the next picture at their own pace by pressing the right arrow. If a participant knew a Spanish word, he/she had to notify the experimenter about it.

Multiple choice task. The multiple choice task had two rounds. During the first round, the words were shown one by one on the screen together with two Spanish words underneath. Participants needed to choose the word that belonged to the picture by clicking on it with the mouse. If they chose correctly, they saw a green circle around the chosen word; if they chose a wrong word, they saw a red circle around the chosen word. Regardless of the correctness of their choice, the correct word was played via the headphones.

In the second round, participants had to make an attempt at naming the word before seeing two options on the screen. The experimenter coded their answers as correct, partially correct or incorrect. After that, two options appeared on the screen, and participants could choose then. The feedback that participants received was the same as in the first round.

Word completion task. In this task, the picture together with the first syllable of the Spanish word were presented on the screen. Participants needed to complete the word and say it out loud. As in the previous task, participants saw a green or a red frame indicating whether they named the word correctly. This feedback was initiated via a button press by the experimenter. Again, the experimenter coded answers as correct, partially correct or incorrect. If the participant pronounced the word correctly, they saw a green frame around the picture; if they pronounced the wrong partially incorrectly/incorrectly, they saw a red frame around the picture. Participants also heard the correct word through the headphones. The stem completion task had two identical rounds.

Writing task. In the writing task, participants saw the picture on the screen and were asked to write down the correct word with a blue pen on a piece of paper (they were provided with a memo paper block). When the word was written down, they could press the right arrow to see how the word was actually spelled. If they made any mistakes, they were asked to take a red pen, correct their mistakes and write the correct word down. After that, participants had to turn the piece of paper over, put it aside and take a new piece of paper for the next word.

Picture naming adaptive task. In the picture naming adaptive task, participants had to name all 23 Spanish words after seeing the picture on the screen. The task consisted of at least two rounds. If any unlearned items remained after two rounds, the task continued until the moment when all items were named correctly at least twice in a row. Regardless of the number of unlearned words, extra rounds always consisted of ten items, thus, sometimes learned words were also repeated. The researcher coded each response as correct, partially correct or incorrect. The participants received feedback about the correctness of their answer on the screen indicated by a green or red frame, and they also heard the correct word through the headphones.

Day 2/Day 3: Spanish Post-test

The Spanish post-test was used to verify which Spanish words were learned after the learning phase. The procedure was similar to the Spanish picture naming adaptive task. However, participants did not receive any feedback. The coding for correctness was the same as in the previous tasks. Importantly, the Spanish final test was conducted twice: for participants in no consolidation group, both tests were conducted on the same day, while for participants in consolidation group, one test was conducted on the second day, and another test was conducted on the third day.

Day 2/Day 3: English Post-test

It was crucial to make the English post-test a surprise task for participants. It followed the second Spanish final test directly, meaning that it occurred on the second day for participants in the no consolidation group and on the third day for participants in the consolidation group. In this task, participants saw all 46 items, that were initially selected for each participant individually by a MATLAB (Version 9.5) (The MathWorks, Inc., 2018) script, and had to name them in English. The experimenter coded their answers as being correct, partially correct or incorrect. Importantly, if synonyms were used (crib for cradle, hatchet for axe etc.), the experimenter coded them as errors. No feedback was given.

Day 2/Day 3: LexTALE

Participants were asked to undergo a lexical test for learners of English (LexTALE; Lemhöfer & Broersma, 2012). In this task, participants saw a word on the screen and had to decide whether it was an existing English word or not. We used the obtained LexTALE scores as a proxy of participants' English vocabulary knowledge (see Table 1).

Finally, the participants were informed that it was the end of the experiment and were asked what they thought the goal of the study was. Sixteen people understood what the goal was, but, as they admitted, this idea occurred to them only after the final English test, which none of the participants expected. Furthermore, the reimbursement was given; participants were paid with VVV gift cards. The amount paid depended on how much time the participant spent in the lab.

The time spent in the lab largely depended on the participant's learning speed and ranged from 1,5 hours to 2,5 hours for all three sessions.

Data Analysis

Coding

The recordings obtained from English pre-test and post-test were used to code participants' answers for accuracy and to measure naming latencies.

As for accuracy, we coded the participant's answers as either correct responses or errors. All trials in which the participant named the picture correctly or corrected herself accurately were considered as correct responses. We counted trials in which the participant named the picture incorrectly, did not give an answer, used a synonym, or corrected herself to a wrong word after saying the correct word as errors.

Naming latencies were determined as the time between the onset of the stimulus and the speech onset with the help of the computer software Praat (Version 6.0.48; Boersma & Weenink, 2019). We set naming latencies to 0 in the following trials: the participant produced a wrong word, made multiple naming attempts, corrected herself, talked, laughed, or coughed. In case the participant used an article, we measured the naming latency at the onset of the article and noted this trial down.

Exclusion criteria

As a first step, one participant who did not learn enough Spanish words during the Spanish learning phase were excluded. The cut-off was 17 learned Spanish words.

Next, we excluded trials with Spanish words that were known to participants before the experiment. The items that were not learned, i.e. not named correctly in the final Spanish test, were also excluded. These trials were not used in both accuracy and naming latencies analyses. Accuracy analysis included all trials that were left after these exclusions. The data of all 54 participants were used in the accuracy analysis.

As for reaction times analysis, in addition to the exclusion mentioned above, we excluded trials with naming latency set to 0 in either pre-test or post-test (described in the section Coding). Moreover, we decided not to include trials in which articles were used. As a result, the data of three participants for who the article trial exclusion resulted in more than 30% trial loss in either interference or no interference condition were removed from the analysis. Thus, the data of 51 participants were used for reaction times analysis.

Modelling

We used R (R Core Team, 2018) and *lme4* (Bates, Maechler, Bolker & Walker, 2015) to perform (generalized) linear mixed effects analyses of the relationship between naming accuracy (the number of errors in the English post-test), condition (interference/no interference) and consolidation group (consolidation/no consolidation) as well as naming latencies (expressed as a difference in reaction times between English pre-test and post-test), condition and consolidation group. Due to the fact that dependent variables were of different nature (accuracy is a binary variable; naming latency is a continuous variable), two separate models were built.

The accuracy data were analysed using a generalized linear mixed effects model of the binomial family, fitted by maximum likelihood estimation, using the logit link function and the optimizer 'bobyqa'. We entered condition (two levels: Interference, No Interference) and consolidation group (two levels: Consolidation, No Consolidation) as fixed effects (with interaction term). The fixed effects were effects-coded. Intercepts for subject and item were used as random effects. Initially, we also included by-subject random slopes for the effect of condition. However, the random slope was perfectly correlated with the intercept for subject, so we left the slope out. All *p*-values were calculated by model comparison, omitting one factor at a time and using chi-square tests.

Naming latencies were analysed using a linear mixed effects model, fitted by restricted maximum likelihood estimation, using Satterhwaite approximation to degrees of freedom and the optimizer 'bobyqa'. We again entered the condition and consolidation group as fixed effects (with interaction term). The fixed effects were effects-coded. Intercept for item as a random effect as well as by-subject random slope for the effect of condition were included. Importantly, we log-transformed the latencies for pre-test and post-test, since the distributions of raw latencies were extremely non-normal. Then we calculated the differences scores, i.e. speed-up in naming latencies from the pre-test to the post-test, and used these for the model. We decided to use the difference scores instead of raw naming latencies, since it is a cleaner measure, which takes into account a priori difference in accessibility between the items at the pre-test. All *p*-values were calculated in the same way as for the accuracy model.

Results

English Pre-test

In the English pre-test, participants from the consolidation group on average could name 67.80% (range: 41.43-89.29; *SD* = 12.40) of the words. The mean of the no consolidation group's performance at the pre-test was 63.20% (range: 33.57-82.86; *SD* = 14.47). An independent samples t-test did not reveal a significant difference in two groups' performance, $t(50.80) = 1.26$, $p = 0.22$.

Spanish Learning

As for the Spanish learning, participants from the consolidation group learned on average 90.82% of the Spanish words (range: 69.57-100; *SD* = 8.64) after completion of the Spanish learning phase, while participants from the no consolidation group learned on average 94.36% (range: 69.57-100; *SD* = 7.42) of the Spanish words. The two groups did not differ in their performance after the Spanish learning, $t(50.83) = -1.62$, $p = 0.11$.

English Post-test

Accuracy

Mean accuracy rates during the English posttest for both the consolidation and the no consolidation group can be inspected in Figure 1.

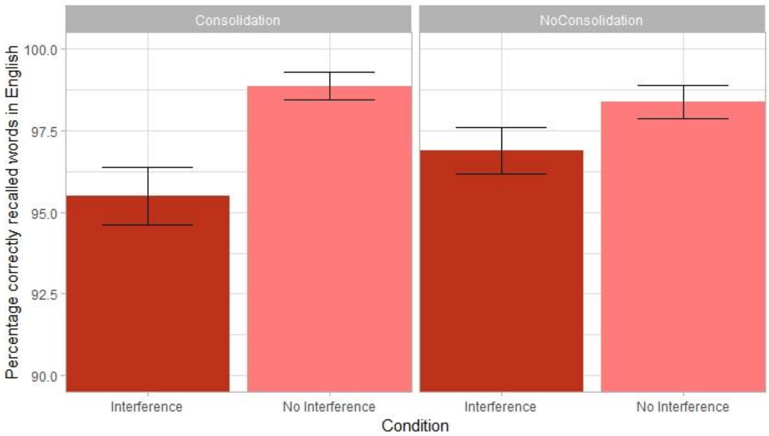

Figure 1. *Mean of correctly recalled English words (in percentages) per group (Consolidation vs. No Consolidation) and condition (Interference vs. No Interference). Error bars indicate 95% CI.*

The estimates of the generalised linear mixed-effects model for accuracy are displayed in Table 3. A significant main effect of interference was found, which means that participants' accuracy differed significantly across the two interference conditions. Specifically, participants made more mistakes in the interference condition than in the no interference condition. However, there was no main effect of consolidation group, meaning that participants in two consolidation groups were equally accurate at the English post-test. Moreover, the interaction term also did not reach significance.

Although the interaction did not come out significant, we decided to conduct post-hoc paired t-tests separately for the two consolidation groups. On the one hand, the t-test showed that participants in the consolidation group differed in their naming accuracy depending on the condition that the word was in (interference vs. no interference), $t(26) = -3.53$, $p = 0.002$, suggesting that they made more mistakes when naming English words interfered with Spanish than non-interfered ones. On the other hand, results of the other t-test indicated that participants in the no consolidation group did not differ in their naming accuracy regardless of the condition, $t(26) = -1.69$, $p = 0.10$, which implies that they were equally accurate at naming words in interference condition and no interference condition. Thus, it can be concluded that even though the interaction did not emerge significant in the full model, the main effect is driven by the consolidation group only, since it is not statistically reliable in the no consolidation group.

Table 3. *Estimates, standard errors, z-values and* p-*values of the generalised linear mixed-effects model for accuracy*.

Fixed effects	**Estimate**	***SE***	***z*-value**	***p*(χ2)**
(Intercept)	-4.18	0.26	-16.04	<.001
Condition	1.09	0.30	3.63	<.001
Consolidation Group	-0.06	0.30	-0.19	0.85
Condition:Consolidation Group	0.79	0.60	1.31	0.19
Random effects	**Variance**	***SD***		
Participant (intercept)	0.00	0.00		
Item (intercept)	1.41	1.19		

Note. SE = Standard Error; *SD* = Standard Deviation.

Naming latencies

Figure 2 displays mean naming latencies the English posttest for both groups. Furthermore, Figure 3 provides the information on mean speed-up in naming latencies from English pre-test to post-test.

The estimates of the linear mixed-effects model for speed-up in naming latencies from English pre-test to post-test are displayed in Table 4. We again found a main effect of interference: relative to the pre-test, participants produced non-interfered words much faster than interfered words at the final English test. However, no main effect of group was observed, indicating that the speed-up was similar across both groups. Finally, there was no interaction between interference and group.

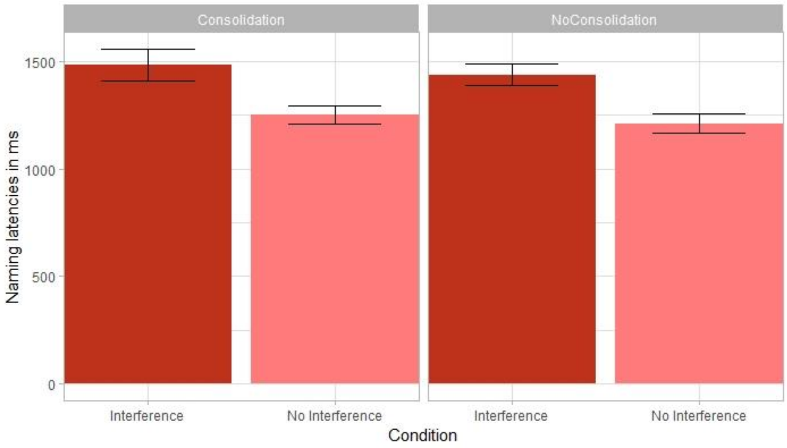

Figure 2. *Mean naming latency of recalled English words (in ms) per group (Consolidation vs. No Consolidation) and condition (Interference vs. No Interference). Error bars indicate 95% CI.*

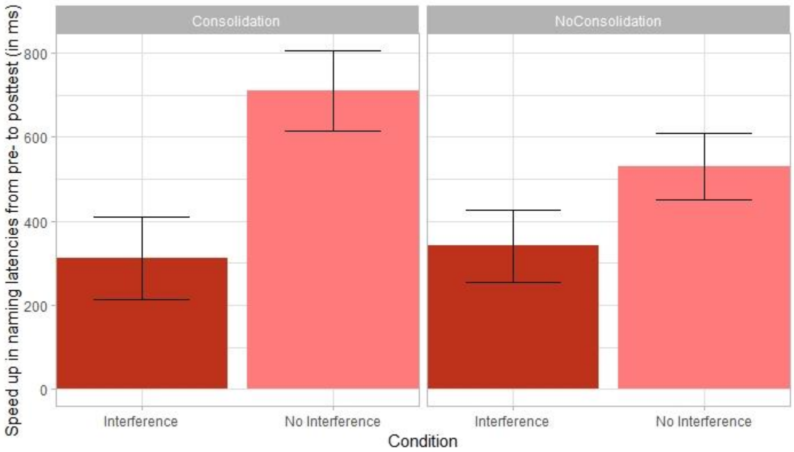

Figure 3. *Mean speed-up in naming latencies from English pre-test to post-test (in ms) per group (Consolidation vs. No Consolidation) and condition (Interference vs. No interference). Error bars indicate 95% CI.*

Table 4. *Estimates, standard errors, t-values and p-values of the linear mixed-effects model for speed-up in naming latencies.*

Fixed effects	**Estimate**	***SE***	***t*-value**	***p*($\chi 2$)**
(Intercept)	0.22	0.03	6.78	<.001
Condition	-0.17	0.03	-6.34	<.001
Consolidation Group	-0.01	0.06	-0.18	0.86
Condition:Consolidation Group	-0.07	0.05	-1.36	0.17
Random effects	**Variance**	***SD***	**Corr.**	
Item (intercept)	0.02	0.16		
Participant (intercept)	0.03	0.18		
Participant (interference)	0.01	0.10	0.39	

Note. *SE* = Standard Error; *SD* = Standard Deviation; *Corr*. = Correlation.

We also performed post-hoc paired t-tests for two consolidation groups. By contrast with accuracy, t-tests revealed significant differences in speed-up in naming latencies across conditions for both of the groups (for consolidation group: $t(25) = -4.96$, $p < 0.001$; for no consolidation group: $t(24) = -3.65$, $p = 0.001$), meaning that participants in two consolidation groups named words from the no interference condition considerably faster than words from the interference condition in the English final test as compared to the English pre-test. Thus, we can say that the main effect is driven by both groups.

Finally, we decided to run a linear mixed-effects model on raw naming latencies as well to make sure that it does not change the outcome. The estimates of this model can be inspected in Table 5. As Table 5 shows, the conclusions from the model for raw naming latencies corroborate the conclusions from the model for speed-up in naming latencies. We found the main effect of interference, which means that participants needed differing amount of time to produce words in two conditions. Namely, participants were slower when naming words which were interfered with Spanish than those that were not. Furthermore, no main effect of group was detected, hence, naming latencies were comparable across the two groups. Lastly, the interaction term again did not come out significant.

Table 5. *Estimates, standard errors, t-values and p-values of the linear mixed-effects model for raw naming latencies.*

Fixed effects	**Estimate**	***SE***	***t*-value**	***p*(*χ2*)**
(Intercept)	7.08	0.03	265.32	<.001
Condition	0.14	0.02	6.17	<.001
Consolidation Group	0.001	0.04	0.03	0.98
Condition:Consolidation Group	-0.01	0.05	-0.26	0.79
Random effects	**Variance**	***SD***	**Corr.**	
Item (intercept)	0.02	0.16		
Participant (intercept)	0.02	0.14		
Participant (interference)	0.01	0.11	0.30	

Note. *SE* = Standard Error; *SD* = Standard Deviation; *Corr*. = Correlation.

Post-hoc paired t-tests were also conducted. T-tests indicated significant differences in naming latencies across conditions for both of the groups (for consolidation group: $t(25) = 3.57$, $p = 0.001$; for no consolidation group: $t(24) = 5.03$, $p < 0.001$), hence, participants in both groups took more time to produce words that they also knew in Spanish than words that they did not learn in Spanish.

Discussion

In this study, we asked whether offline consolidation, including a night's sleep, modulates the degree to which the novel words in a new foreign language (L3) interfere with the translation equivalents in an already known foreign language (L2), and further, whether this interference can induce forgetting in an already known L2. Dutch monolinguals with good knowledge of English (L2) and no or little knowledge of Spanish (L3) first did a picture naming task in English. On the basis of the participant's right answers, a set of English words for the further experiment was generated. After that, participants learned Spanish translation equivalents of half of the words from this set. Finally, they did another picture naming task in English, in which they had to name all the words from the chosen set. Importantly, participants were divided into two groups: one group, referred to as no consolidation group, learned Spanish words and performed the final English test on the same day. Another group, referred to as consolidation group, was taught Spanish words on one day and did the final English task one day later. That way, newly learned Spanish words were given time to become consolidated.

Our hypothesis was that learning of words in the L3 would lead to interference in the participants' L2, which could show in higher naming latencies as well as lower naming accuracy for the interfered words. Based on the previous work by Mickan et al. (in preparation), we expected to observe the interference effects at least in naming latencies in both consolidation groups. The hypothesis was confirmed. Specifically, after the interference in the form of learning of new Spanish words was introduced, participants in both groups needed significantly more time to name the corresponding words in English, as compared to the words that were not interfered with Spanish. Furthermore, we found a main effect in naming accuracy across both groups, meaning that participants made significantly more mistakes in words, that were learned in Spanish, at the final English test in comparison with the English pre-test. Importantly though, when we tested for main effects in naming accuracy in the two groups separately via t-tests, we observed that the interference effect in naming accuracy was reliable only in the consolidation group.

These results are in line with previous research, and they suggest that foreign languages that the person knows influence each other bidirectionally and are able to interfere with each other, which, in turn, results in an impairment at retrieval (Mickan et al., in preparation; Mickan et al., submitted; Isurin & McDonald, 2001). Similarly to the previous studies, in the present study the interference effects were manifested in both higher naming latencies and lower naming accuracy for the interfered English words. A possible explanation for the observed effects could be the following: novel Spanish words, connected to English words via a shared concept, entered the lexical competition at the moment when participants were asked to name a picture in English at the final test and, thus, impaired the access to the target English words. Consequently, participants made more mistakes in their English productions as well as took longer to produce the words in English. Furthermore, participants had to repeatedly retrieve Spanish words during the Spanish learning stage, which resulted in strengthening of the memories for these words, but damaged the access to non-retrieved English translation equivalents. Importantly, the suggested explanation for our findings closely resembles the way RIF is described (Anderson et al., 1994), hence, it could be the mechanism underlying language forgetting.

Moreover, we were wondering whether the interference effects we discovered would become stronger if Spanish words, that were learned in the experiment, would be consolidated overnight. If this was the case, we could observe either of the following patterns: a stronger effect in naming latencies in the consolidation group, a stronger accuracy effect in the consolidation group (even when it is present in both groups), or an accuracy effect in the consolidation group when it is absent in the other group. Ultimately, our data followed the last pattern. To be precise, in regard to naming latencies, we did not find a greater interference effect in the consolidation group. In regard to naming accuracy, t-tests showed that there was a statistically reliable effect of interference in the consolidation group, but not in the no consolidation group, as was mentioned above, which pointed towards a role for consolidation in driving interference. It is essential to note though that this was not confirmed in a direct statistical comparison of the groups (i.e. no interaction effect), hence, not allowing us to make for any firm conclusions as to the role of consolidation. Nevertheless, the results tentatively suggest that consolidation of newly learned words in L3 causes a more significant interference effect in naming accuracy, but not in naming latencies. This finding is partly in line with the earlier work on consolidation effects after a night of sleep. The accuracy effect observed in the consolidation group and not in the no consolidation group is a manifestation of a close association between overnight consolidation of novel words and the emergence of lexical competition effects leading to greater interference, which has been demonstrated previously (for an overview see Davis & Gaskell, 2009). However, in relation to naming latencies, it has been claimed that only words consolidated overnight are able to significantly slow down the retrieval of competitor words (Dumay & Gaskell, 2007), which was not the case in the present study.

The absence of a difference in the interference magnitude in naming latencies between the two groups and a rather weak interference effect in accuracy in the consolidation group, which could not be detected via a direct statistical comparison, can be explained as follows. As mentioned in Introduction, there exists evidence of rapid systems-level consolidation, suggesting that a single period of sleep is enough for changes in the way novel representations are stored, i.e. the transition from the hippocampus to neocortex (e.g. Dumay & Gaskell, 2007; Tamminen, Davis, & Rastle, 2015). Reactivation of hippocampal memory traces during sleep indeed promotes the increase in their strength and robustness, and that can account for the lexical competition and related phenomena found after one night's sleep. However, some researchers suggest consolidation of new memories and their full integration into neocortex takes considerably longer (Gaskell & Dumay, 2003; Kumaran & McClelland, 2012; Kumaran, Hassabis, & McClelland, 2016). For instance, Gaskell & Dumay (2003) observed the lexicalisation effects only one week after the novel words were learned. Hence, it might be the case that if the novel Spanish words in our study were given more time to stabilise, a significant difference in the interference magnitude in naming latencies between the groups would arise. Moreover, we would probably observe a more robust effect in naming accuracy in the consolidation group, allowing us to make more solid conclusion in regard to the role of consolidation.

All in all, the strength of the present study is its conformity with previous findings in regard to the interference effects. We found that language interference slowed down subsequent naming and led to more errors in Spanish word productions in comparison to the no interference condition, thus, corroborating the previous findings (Mickan et al., submitted; Mickan et al., in preparation; Isurin & McDonald, 2001; Levy et al., 2007). Furthermore, the theoretical framework and the design of this study were broadly similar to the study of Mickan et al. (in preparation). The set-up of all the Spanish learning tasks was identical, and the experimental stimuli broadly overlapped. Thus, our study provided further support for the claimed made by Mickan et al. (in preparation), indicating that new learning leads to retrieval difficulties in previously learned foreign languages, which can be explained via between-language competition. Crucially, as opposed to the study by Mickan et al. (in preparation), we managed to show the actual language forgetting, i.e. the effect in accuracy, so it can be said the present study simulated the precursors of foreign language attrition in the laboratory setting.

Importantly, this study also has a few limitations. Namely, one of the limitations is that all our participants were fluent speakers of English. Hence, the findings can only be applied to highly proficient speakers of L2. Future research should investigate the effects of consolidation of newly learned L3 words on an L2 in less proficient L2 speakers. Even larger effects of interference are to be predicted, since the less L2 is established in the brain, the more susceptible it is to interference and deterioration (Mehotcheva, 2010; Schmid & Mehotcheva, 2012). Another limitation is that the consolidation period for an L3 in our study was only one night. As has been suggested above, consolidation is a gradual process, that is likely to take longer than a single period of sleep (Kumaran et al., 2016). More research is needed to explore the consolidation period of an L3 and its effects on an L2 further, since so far we did not find clear evidence of a modulation of interference by consolidation. The expectation is that a more consolidated/entrenched L3 would be a stronger interferer for an L2.

In conclusion, our study has demonstrated that the learning of a new foreign language results in interference in and forgetting of a previously known foreign language, as seen in higher naming latencies and lower naming accuracy. What is more, our results imply that overnight consolidation of newly learned words in an L3 might lead to a stronger interference effect in naming accuracy, but not naming latencies. These findings broaden our understanding of how people forget languages and also, more generally, about how foreign languages interact with one another. Combined with a lot more research on foreign language attrition, second language acquisition, and cognitive neuroscience, it may lead to the development of a certain educational program or learning strategies that will help people learn languages so that they are not subsequently forgotten.

References

Anderson M. C., & Spellman B. A. (1995). On the status of inhibitory mechanisms in cognition: memory retrieval as a model case. *Psychological Review*, 102, 68–100

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1063- 1087

Anderson, M.C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, 49, 415–445.

Bardovi-Harlig, K., & Stringer, D. (2010). Variables in Second Language Attrition. Advancing the State of the Art. *Studies in Second Language Acquisition*, 2010(32), 1-45. doi: 10.1017/S0272263109990246

Bates, D., Maechler, M., Bolker, B. & Walker S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi: 10.18637/jss.v067.i01.

Bergmann, C., Sprenger, S. A., & Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta psychologica*, 161, 25-35. doi: 10.1016/j.actpsy.2015.07.015

Boersma, P. & Weenink, D. (2019) Praat: doing phonetics by computer (Version 6.0.48) [Computer software]. Retrieved from: http://www.praat.org/

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, 8(3), 278–302.

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psych*ology. General, 143(6), 2296–2303.

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

Davis, M.H., & Gaskell, M.G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society B*, 364, 3773–3800.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In Kroll & de Groot (eds.), pp. 179–201.

Dijkstra, T., Grainger, J. & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, 41, 496–518.

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

Ecke, P. (2004). Language attrition and theories of forgetting: A cross-disciplinary review. International *Journal of Bilingualism*, 8(3), 321-354.

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. Nature Reviews Neuroscience, 6(2), 119–130.

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

Isurin, L., & Mcdonald, J. L. (2001). Retroactive interference from translation equivalents: Implications for first language forgetting. *Memory & Cognition*, 29(2), 312-319. doi: 10.3758/BF03194925

Kaushanskaya, M. & Marian, V. (2007). Bilingual language processing and interference in bilinguals: Evidence from eye tracking and picture naming. *Language Learning*, 57, 119–163

Keppel, G. (1968). Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*, 172 – 213. Englewood Cliffs, NJ: Prentice-Hall.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences*, 20(7), 512–534. doi: 10.1016/j.tics.2016.05.004

Lemhoefer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44, 325–343.

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second-language acquisition. *Psychological Science*, 18(1), 29-34. doi: 10.1111/j.1467-9280.2007.01844.x

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. Journal of Experimental Psychology. *Learning, Memory, and Cognition*, 39(2), 608–622.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, 92, 57-78. doi: 10.1016/j.jml.2016.04.001

Marr, D. (1970). A theory for cerebral neocortex. *Proceedings of the Royal Society Series B*, 176(1043), 161–234.

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.

McClosky, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation*, 109–165. New York: Academic Press.

Mehotcheva, T.H. (2010). After the fiesta is over: Foreign language attrition of Spanish in Dutch and German Erasmus students. Doctoral dissertation, University of Groningen. GRODIL Series (86)/RECERCAT, Pompeu Fabra University.

Mickan, A., McQueen, J.M., K. Lemhöfer (2019). The role of between-language competition in foreign language attrition. Manuscript submitted for publication.

Nader K., & Hardt, O. (2009). A single standard for memory: the case for reconsolidation. *Nature Reviews Neuroscience*, 10, 224–234, 2009.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition*, 1, 19 – 40.

R Core Team (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database [webcelex]. Philadelphia, PA: University of Pennsylvania Linguistic Data Consortium.

Schmid, M. S., & Mehotcheva, T. (2012). Foreign language attrition. *Dutch Journal of Applied Linguistics*, 1(1), 102-124. doi: 10.1075/dujal.1.1.08sch

Spear, N. Retrieval of memory in animals. (1973). *Psychological Review*, 80, 163–194.

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., Davis, M.H. and Rastle, K. (2015). From Specific Examples to General Knowledge in Language Learning. *Cognitive Psychology*, 79, 1-39. doi: 10.1016/j.cogpsych.2015.03.003

The MathWorks, Inc. (2018). MATLAB (Version 9.5) [Computer software]. Retrieved from: https://nl.mathworks.com/products/matlab.html

Appendix

140 stimuli used in the English pre-test and their Spanish and Dutch translations

Spanish	Dutch	English
enchufe	stopcontact	socket
marco	lijst	frame
escarabajo	kever	beetle
taburete	kruk	stool
alambrada	hek	fence
tiza	krijtje	chalk
grifo	kraan	tap
regla	liniaal	ruler
baúl	kist	chest
fregona	dweil	mop
pastel	taart	cake
raíz	wortel	root
gorro	muts	hat
grapadora	nietmachine	stapler
taladro	boor	drill
tapa	deksel	lid
pañal	luier	diaper
rama	tak	branch
cerillo	lucifer	match
tobogán	glijbaan	slide
arbusto	struik	bush
lata	blik	can
abrigo	jas	coat
escoba	bezem	broom
sierra	zaag	saw
canasta	mand	basket
ladrillo	baksteen	brick
tarro	pot	jar
silbato	fluit	whistle
dado	dobbelsteen	dice
hacha	bijl	axe
plancha	strijkijzer	iron
hormiga	mier	ant
manta	deken	blanket
aguja	naald	needle
arco	boog	bow
cinturón	riem	belt
cuenco	kom	bowl
guante	handschoen	glove
loro	papegaai	parrot
manga	mouw	sleeve
pala	schep	shovel
melocotón	perzik	peach
paja	rietje	straw
tortuga	schildpad	turtle

uvas	druif	grape
columpio	schommel	swing
jaula	kooi	cage
clavo	spijker	nail
moneda	munt	coin
mosca	vlieg	fly
cadena	ketting	chain
bala	kogel	bullet
diente	tand	tooth
hoja	blad	leaf
caracol	slak	snail
concha	schelp	shell
pulsera	armband	bracelet
burro	ezel	donkey
cubo	emmer	bucket
bufanda	sjaal	scarf
flecha	pijl	arrow
corbata	stropdas	tie
mechero	aansteker	lighter
rana	kikker	frog
soga	touw	rope
paraguas	paraplu	umbrella
almohada	kussen	pillow
ardilla	eekhoorn	squirrel
lápiz	potlood	pencil
naranja	sinaasappel	orange
nube	wolk	cloud
ajo	knoflook	garlic
cremallera	rits	zipper
mono	aap	monkey
sello	postzegel	stamp
sujetador	beha	bra
toalla	handdoek	towel
cartera	portemonnee	wallet
hueso	bot	bone
traje	pak	suit
muñeca	pop	doll
salchicha	worst	sausage
falda	rok	skirt
madera	hout	wood
ventana	raam	window
pájaro	vogel	bird
pluma	veer	feather
araña	spin	spider
árbol	boom	tree
cacahuete	pinda	peanut
cebolla	ui	onion
cuchillo	mes	knife
ola	golf	wave

reloj	horloge	watch
vela	kaars	candle
zanahoria	wortel	carrot
bigote	snor	moustache
cuchara	lepel	spoon
espejo	spiegel	mirror
llave	sleutel	key
silla	stoel	chair
tiburón	haai	shark
ala	vleugel	wing
avestruz	struisvogel	ostrich
batidor	garde	whisk
caña	hengel	rod
cerdo	varken	pig
colador	zeef	strainer
embudo	trechter	funnel
jabón	zeep	soap
llanta	velg	rim
machacador	stamper	masher
manopla	want	mitten
mapache	wasbeer	raccoon
oruga	rups	caterpillar
pavo	kalkoen	turkey
peluca	pruik	wig
pepinillo	augurk	pickle
puño	vuist	fist
queso	kaas	cheese
rallador	rasp	grater
regalo	cadeau	gift
taza	kopje	mug
tornillo	schroef	screw
tuerca	moer	nut
yedra	klimop	ivy
cuna	wieg	cradle
paloma	duif	pigeon
pierna	been	leg
serpiente	slang	snake
vestido	jurk	dress
caja	doos	box
huevo	ei	egg
pato	eend	duck
caballo	paard	horse
coche	auto	car
bastón	stok	cane
canica	knikker	marble
abanico	waaier	fan