

Effects of Repeated Word Retrieval on the Use of Keyword Mediators

Shifting from Mediated to Direct Retrieval Predicts Better
Learning Outcomes

Name: Mirte Dikmans

Student number: 4225872

Date: 03-03-2018

Master: General Linguistics

Supervisor: Jetske Klatter

Second supervisor: Gesa van den Broek

Table of contents

Abstract	1
Introduction	1
Keyword method	2
Retrieval practice	4
Shift in memory retrieval.....	5
Think aloud method.....	8
Present study.....	10
Method	13
Participants	13
Overview of the experiment	13
Design.....	13
Materials	13
Procedure	18
Data preparation	21
Analyses.....	23
Results	24
RQ1: When do learners shift from mediated to direct retrieval?	25
RQ2: Do shift differences between learners predict differences in performance later on?..	29
RQ3: Do shift differences between words predict differences in performance later on?.....	31
Discussion	34
Implications	36
Limitations.....	37
Additional observations.....	39
Conclusion.....	41
Literature	43
Appendix	46

Abstract

The present study investigated how the use of keyword mediators changes over the course of repeated retrieval practice and how these changes relate to the retention of foreign vocabulary by means of the think aloud method. Keyword mediators are used in vocabulary learning to associate novel word forms to their meaning (e.g., *barua=letter*), by selecting a keyword that resembles the word form (e.g., *bar*) and linking it to the word meaning by means of a mental image (e.g., *she writes a letter in a bar*). 30 students encoded novel words using experimenter-provided keywords and then repeatedly retrieved the word meanings while thinking aloud. The think aloud protocols demonstrate a decrease in mediator for 21.6% of the experimental words. These shifts from mediated to direct retrieval are explained to be a function of repeated practice rather than time. The findings show that the average moment of shift appeared after 8.27 retrievals. Further, no correlation was found between the mediator use and the test performance one week after practice on learner level. On word level however, the mediator use during practice could predict both the receptive and productive recall on the test. The establishment of a direct link between the target form and meaning seems to enhance the retention of words that were initially learned with keyword mediators.

Introduction

Having extensive knowledge of vocabulary is essential in order to successfully use and understand a language (Schmitt, 2008). Learning the vocabulary of a foreign language is therefore an important part of foreign language acquisition. Basically, the process of vocabulary acquisition comes down to mapping a new, unfamiliar sound onto an already existing semantic representation in the memory (Ellis, & Beaton, 1993a). The storage and mapping of foreign vocabulary items in the long-term memory is enhanced by features in the mental lexicon that already exist. The mental lexicon contains all the information about the words that are known by an individual, containing among others phonological, orthographic and semantic features. Together, all features form the representation of a word in the mental lexicon (Emmorey, & Fromkin, 1988). The semantic representations in the mental lexicon are related in a semantic network consisting of nodes (i.e., word representations) and connections between these nodes (Collins and Loftus, 1975). The distances between the nodes reflect the semantical distance between the word representations and the strength of the connections between these nodes reflect the relation between the word representations (Lowe, 1997). By mapping new information onto existing features in the mental lexicon, the connection between the target word and its meaning becomes stronger and better represented in the long-term memory. This mapping process is also present when learners use already familiar words to learn new vocabulary items. In this way learners map new, foreign words onto already existing word representations in the memory (Schmitt, 2000).

Learners can make use of many different strategies to optimize the process of vocabulary learning in a foreign language. However, not every strategy is equally successful. Dunlosky et al. (2013) discussed and evaluated ten different learning techniques, among which practice tests, summarizing, highlighting, rereading, and mnemonic strategies. It was found that some frequently used learning techniques, like highlighting and rereading, are often less useful and effective than other, less frequently used techniques, like distributed practice and practice testing. Importantly, they argue that a considerable number of learners could easily switch to more efficient learning techniques that expectedly enhance foreign language learning and lead

to better retention. Possibly, the most effective and useful learning strategy might be practice testing. Practice testing is a learning strategy in which learners use tests for learning purposes rather than for evaluation purposes, in order to improve their knowledge. A different strategy that could be useful for foreign vocabulary learning is the use of keyword mediators. When using keywords for the acquisition of novel vocabulary, learners use mediators to associate and learn word pairs based on acoustic similarity, which could be a very effective learning strategy provided that the foreign vocabulary materials and the learning circumstances are suitable for mediator use (Atkinson, 1975).

Keyword method

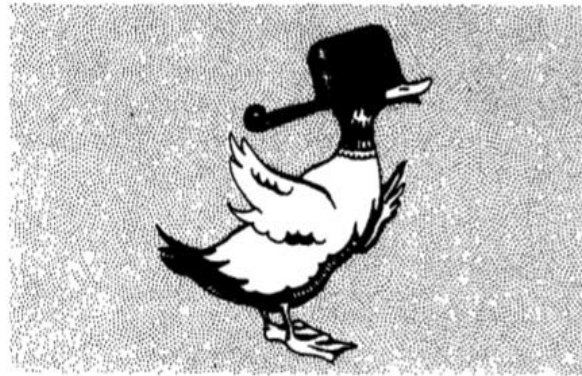
General description

Learners can benefit from the use of mnemonic keywords when learning the vocabulary of a foreign language (e.g., Atkinson, 1975; Fritz et al., 2007). Keywords are words that are already familiar to the learner and are acoustically similar to a novel, foreign word that the learner intends to acquire. Because of the similarities between the keyword and the novel word, the keyword functions as a mediator between the form and the meaning of the foreign word, and therefore facilitates the establishment of a link between target word and its corresponding translation. This helps the learner to connect the foreign word form with its meaning when learning the vocabulary of a foreign language (Atkinson, 1975).

Three steps need to be performed in order to create a link between the form and the meaning of the target word when using the keyword method (Hulstijn, 1997). The first step is to select a familiar word, for instance in the native language of the learner, that is acoustically similar to the target word. The second step is to associate this familiar keyword with the meaning of the foreign target word. The learner could do so by creating an interaction between the keyword and the translation. The third and last step is to form a vivid mental image of this interactive association between the keyword and its translation.

This learning process is best explained by means of an example. An example of a possible mental image when learning the Spanish word *pato* is depicted in Figure 1 (Atkinson, 1975, p. 822). Learners could choose the English word *pot* as a keyword for *pato* based on acoustic similarity. Next the learner should create a vivid visualization of an interaction between the keyword and the meaning of the foreign word, like imagining a *duck* with a *pot* on its head. Preferably this association should be a bit strange and uncommon, so that the association will be remembered best. This mental image later comes to mind when the learner is presented with the Spanish word *pato*. In this way the retrieval of the word meaning will be enhanced (Atkinson, 1975).

Using mental imagery to create a link between the target word and its meaning could be very beneficial for vocabulary learning for many different reasons. For instance, mental imagery makes the learning task more interesting and therefore the learners get more motivated. Further, mental imagery could enhance the distinctiveness between the correct meaning of a word and the incorrect competitors, which increases the accurate retention of the newly learned information and decreases the recall of wrong combinations (Bower, 1972). Besides, creating and visualizing a mental image requires the learners to actively integrate and combine the meaning of both the keyword and the translation, which also enhances the differentiation between the correct translation and the competitors (Dunlosky et al., 2013). Additionally, mental imagery makes the combination of the keyword and the target more logical and meaningful to the learners, since they actually visualize an interaction between the two parts in an active way themselves (Karpicke & Smith, 2012).



PATO — pot — DUCK

Figure 1. An example of a mental image for the Spanish word *pato* and the corresponding translation. Reprinted from “Mnemonotechnics in Second-Language Learning” by R. C. Atkinson, 1975, *American psychologist*, 30(8), p. 822.

Given these points, mental imagery seems to strengthen the connection between the foreign word and its meaning and enhance the discrimination between the correct meaning of a word and the incorrect competitors. Later, when retrieving words that were learned using the keyword method, the mental image pops into mind when the foreign word is presented and this helps the learner to remember the translation of the word. Altogether, this forming of a mental image is crucial and it might even be the most important step when applying the keyword method for vocabulary learning.

Effectiveness

The keyword method can be a beneficial vocabulary learning strategy. Atkinson (1975) found evidence supporting the effectiveness of the keyword method in both a study on Russian and Spanish vocabulary learning. First, in the study on Russian vocabulary the participants were either instructed to use the keyword method or they did not receive any instructions about a learning technique at all. The results showed that the participants who used the keyword method outperformed the control group, as they recalled 72% of the words correctly compared to 46% in the control group. Even stronger evidence was found in the Spanish vocabulary study. This study was similar to the Russian vocabulary study, except that the participants in the control group were instructed to use rote rehearsal, in which learners frequently repeat the novel information in their minds, to learn the words. This time the difference in correctly recalled words between the two groups was even bigger, as the keyword group recalled 88% of the words correctly and the control group recalled only 28% correctly. This indicates that the keyword method can be a very effective method to learn the vocabulary of a foreign language, especially compared to the rote rehearsal learning method.

The effects of the keyword method on the longer term seem to be less robust. Dunlosky et al. (2013) argue that accelerated forgetting might appear and that the retention of the words decreases over time. However, a study with Iranian learners of English found that the keyword method positively affected the performance of the learners on both an immediate test and a delayed test after two weeks. The keyword group again strongly outperformed the control group and the forgetting of the words in two weeks' time was much less significant for the keyword group than for the control group (Sayyadi, & Javadi, 2014). Furthermore, positive effects of the keyword method have been found on multiple language levels (Piribabadi, & Rahmany, 2014)

and in various learner and learning circumstances (Safa, & Hamzavi, 2013), indicating that the positive effect of the keyword method seem to hold in many different learning situations.

The mediator effectiveness hypothesis (Pyc, & Rawson, 2010) might provide an explanation for the effectiveness of the keyword mediators. This hypothesis states that keywords are effective as a function of two characteristics: the mediator can easily be retrieved from memory and subsequently the mediator elicits the target information from memory. Thus, the mediators are effective when they are already represented in the mental lexicon, since they can be retrieved from memory with ease. Because of the acoustic similarities between the mediator and the target word, the retrieval of the meaning of the target word is enhanced by means of an interactive mental image. Thus, the easily retrieved mediator enhances the retrieval of the mental image and this mental image enhances the retrieval of the target meaning, which makes the keyword method an effective strategy for vocabulary learning.

Although limited evidence is available regarding the effectiveness of keyword mediators on the long term, altogether there is agreement about the effectiveness of the keyword method. This method could get even more effective when combined with another learning strategy, namely with retrieval practice. Fritz et al. (2007) compared restudying, retrieval practice, and the keyword method and found that both retrieval practice and the keyword mnemonic strategy were comparable in effectiveness and both strongly outperformed the restudying strategy. Furthermore, when these two successful learning strategies were combined, the performance of the learners increased even more. Therefore, it is concluded that the combination of retrieval practice and the mnemonic learning strategy seems to be very effective and that these two learning strategies could successfully complement one another.

The broad positive effects of the keyword method for foreign vocabulary learning support the effectiveness of this learning strategy, especially when this method is combined with retrieval practice. Before making reliable predictions about this combination, the following section will thoroughly discuss the retrieval practice learning strategy.

Retrieval practice

Retrieval practice could be advantageous for the learning of the vocabulary of a foreign language (e.g., Karpicke, & Roediger, 2008; Roediger, & Butler, 2011; Karpicke, & Smith, 2012; van den Broek, Segers, Takashima, & Verhoeven, 2014). When using this learning strategy, learners repeatedly and actively recall novel information from memory, which leads to the reinforcement of the memory representation of the foreign word (Nation, 2001). Actively retrieving the newly learned word from memory enhances the later retention of the foreign word. This positive effect of retrieval practice is also known as the testing effect. The testing effect is the advantage of repeated retrieval practice when retaining new information and is especially effective when it is combined with feedback on the performance during practice (Roediger, & Butler, 2011). Apparently, testing could be used not only to test the knowledge on newly learned information, but also to actually learn new information.

Like the keyword method, retrieval practice is considerably more effective than restudy activities, during which words are studied together with their translation (Karpicke, & Roediger, 2008; Roediger, & Butler, 2011; Karpicke, & Smith, 2012; van den Broek, Segers, Takashima, & Verhoeven, 2014). For instance, Karpicke and Roediger (2008) investigated the effectiveness of the testing effect in a study on word learning. They tested participants who were learning English-Swahili word pairs using either retrieval practice or restudying. When the participants were tested on their performance on the same words one week later, the retention of the words was considerably better for participants who had used retrieval practice to learn the words than for participants who had used restudying. This illustrates that words are retained better in case they are learned using the retrieval practice technique instead of the restudying technique.

More evidence for the effectiveness of retrieval practice comes from brain imaging studies, in which positive testing effects on the accuracy of word retention were found (e.g., van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013; Hashimoto, Usui, Taira, & Kojima, 2011; Eriksson, Kalpouzos, & Nyberg, 2011). An fMRI study on the neural processes that underlie the testing effect found that a number of brain areas were more activated during practice testing than during restudying, such as areas regarding semantic processing and memory strengthening (van den Broek et al., 2013). In accordance with the keyword method, the retrieval practice method also seems to strengthen the underlying memory representation of newly learned information and it seems to be a more effortful learning strategy than restudying.

Interestingly, many studies found that the positive testing effects are particularly visible on delayed tests that are performed some time after learning (Carpenter, Pashler, Wixted, & Vul, 2008; Kornell, Bjork, & Garcia, 2011; Wheeler, Ewers, & Buonanno, 2003). However, the positive testing effects are found to be present immediately after practice, but these effects are only visible in case feedback is provided during learning or in case the retrieval success is high. Also, the positive testing effects regarding the retrieval time of the target were directly visible in the quick reaction times of the learners when they were retrieving the newly learned information (van den Broek et al., 2014).

Thus, the positive effects of retrieval practice seem to be compelling, both immediately after practice and on a longer term. Retrieval practice enhances the strength of the newly learned associations between the target words and the corresponding translations. Furthermore, retrieval practice seems to facilitate the memory retrieval process. This facilitation supposedly arises because the discrimination of the correct translation and the incorrect competitors improves over the course of practice. Like the keyword method, retrieval practice seems to increase the distinctiveness between the correct word meaning and the incorrect competitors (Anderson, Bjork, & Bjork, 1994).

Both the keyword method and retrieval practice seem to be very effective strategies for foreign vocabulary learning. A combination of the two might be even more effective. The combination between actively retrieving information from memory together with the mental imagery of the keyword method might be a successful way to strengthen memory representations and enlarge the distinctiveness of the correct translation from its competitors. Based on the evidence regarding the effectiveness of both learning strategies, the present study will combine these two strategies in order to create an optimal learning situation in which the use of the keywords will be investigated.

Shift in memory retrieval

Different models on memory retrieval

When using the keyword method for vocabulary acquisition, learners first select an acoustically similar word that they already know, subsequently this word is linked to the target word in the mental lexicon, based on phonological similarities (Atkinson, 1975). The keyword then helps to establish a link between the form and the meaning of the target word. Repeated retrieval practice could enhance the strength of the memory representation, since the learners actively retrieve information from memory for multiple times (Nation, 2001). This strengthens the link between the target and the correct corresponding translation, that had initially been established by means of keyword mediators.

There are several models that intent to explain how the word form, the keyword and the word meaning are represented in the long-term storage of word representations in the memory, which we here refer to as the mental lexicon (Emmorey, & Fromkin, 1988), and what happens when these words are retrieved. One plausible hypothesis is addressed in the direct model (Adams, & McIntyre, 1967). According to this model, the mediator that is used to link the target

word to its meaning does not become a part of the underlying memory structure. Instead, the target and the corresponding meaning are directly linked and the mediator serves merely as a learning context. Hence, the connection between the target form and meaning is not interrupted by the keyword, but the keyword only helps to establish and remember the direct link. So according to this the direct model, it is possible to retrieve the meaning without using the keyword mediator.

Another plausible hypothesis is addressed in the mediation model (Adams, & McIntyre, 1967). This model suggests that the mediator becomes a significant part in the underlying memory structure in a way that it is only possible to get to the meaning of a word through the mediator. The link between the target and its meaning is mediated, because the retrieval of the word meaning goes through the keyword mediator, making it impossible for the word meaning to be retrieved without using the keyword mediator.

Even though the direct model and the mediation model differ considerably in their claims about how information is retrieved from memory, the models may both be correct in case the nature of the memory structures change depending on the strength of the newly learned connections. The connections in the mental lexicon, as well as the corresponding strengths, are changeable and therefore the keyword mediators may not always have the same influence on the link between the form and meaning of the target word (Hulstijn, 1997). The link between the target word and its meaning might be temporarily mediated until a direct link has been established. After a certain amount of repetitions, a direct link between the foreign word and its translation is likely to arise, since repeatedly rehearsing the retrieval of the translation leads to quick retrieval of the word meanings without the necessity of retrieving the keyword mediators first (Atkinson, 1975). Thus, after the initially mediated phase, the learners would be able to directly retrieve the meaning of the foreign word from memory without using the keyword mediator (Atkinson, & Wescourt, 1974; Atkinson, 1975; Hulstijn, 1997).

Another account is the covert mediation model (Crutcher, & Ericsson, 2000). This model posits that after repeated practice the explicit recall of keywords in the short-term memory disappears, but implicit recall of the keywords in the long-term memory occurs instead. Hence, the keywords are absent in the conscious working memory, but are still activated as part of the underlying memory representation. This model suggests that the retrieval does not switch to direct retrieval after repeated practice but remains mediated in the underlying memory structure. However, the awareness about the mediation varies, depending on the amount of repeated practice. This model connects to several holistic strengthening models (Pirolli, & Anderson, 1985; Anderson, Fincham, & Douglass, 1997), which propose that retrieval practice only strengthens the already existing connections to access memory representations, instead of fundamentally changing the underlying memory structure. This would mean that the initially mediated connections remain mediated, but the strengths of the connections between the target word and the keyword will vary after practice.

Finally, the identical elements (IE) model (Rickard, & Bajic, 2006) supposes that over the course of retrieval practice, independent memory representations emerge, depending on the specific combinations of a cue and a target that are repeatedly retrieved from memory. This model suggests that the initially learned combinations of the foreign word, keyword and translation will not be strengthened as a whole, but mainly the connection between the foreign word and the translation will be strengthened, since this is the combination that is repeatedly retrieved from memory. Thus, this model connects to the direct model and proposes that the underlying mediated connections fundamentally change into direct connections that enhance the retrieval of the target meaning from memory over the course of practice.

Determining the best model

Thus, several models propose explanations for the structure of word representations in memory when vocabulary items are learned with keyword mediators, and the effect that memory retrieval has on these representations. The following section aims to determine the most plausible model, which will function as the fundamental model for the questions and predictions of the present study.

The composition of the memory representation has been investigated by Crutcher and Ericsson (2000). The participants learned and recalled novel English-Spanish word pairs and reported their thoughts immediately after each trial by means of verbal reports. These verbal reports reflected the thoughts that the participants had during the retrieval task. The reports showed that the vast majority of the keywords were still activated after moderate practice (about 6 repetitions), since there were still references to the keywords for most of the items. After very extended practice however, when the words were repeated about 80 times, there were barely any references to the keywords in the verbal reports. This finding indicates that the keywords were no longer used after extended retrieval, which suggests that a shift in memory retrieval might appear after a sufficient amount of repetitions.

The results of this study are explained by means of the covert mediation model, claiming that the retrieval of the target meaning does not require explicit activation of the keyword in the working memory, but instead the retrieval of the target meaning in the long-term memory remains mediated by the keyword (Crutcher, & Ericsson, 2000). This idea of covert mediation was explicitly rejected by two studies, that propose an alternative explanation for these results. Both studies argue that these results support the direct model, which assumes that the structure of word representations changes from a memory representation with mediated connections to a memory representation with direct connections after repeated practice.

First, Rickard and Bajic (2003) reject the covert mediation model and used a non-linguistic task to demonstrate a shift from initially mediated memory retrieval to direct memory retrieval. Instead of using keywords to learn vocabulary, the participants had to learn associations between colours, digits, and letters. First, they had to learn the connection between the colours and the letters, then between the letters and the digits, and finally they were tested on the connections between the colours and the digits. In addition, an interference phase was added which included novel mediators. However, these novel mediators did not interfere in the retrieval of the target. This finding indicates that the connection between the cue and its target, which was initially learned by means of a specific mediator, was not mediated anymore. This finding indicates that a shift from initially mediated retrieval to direct retrieval had occurred.

Second, a continuation of mediated retrieval, as is proposed by the (covert) mediation models, is rejected by Kole and Healy (2013). In their study they aimed to determine an accurate model to describe the retrieval process of words that are learned using the keyword method. Specifically, they investigated whether a shift from mediated retrieval to direct retrieval appears, by means of a mediated priming method. The participants first learned and practiced foreign words using the keywords method and after each trial they performed a lexical decision task in a semantic priming experiment. In this priming task, stimuli were used that were either semantically related to the keyword mediator or not. Participants had to decide on the lexicality of these words. It was expected that in case the prior retrieval was still mediated, so if the keyword mediator had just been activated, the participants would be primed in the lexical decision task. Hence, the participants were expected to recognize semantically related words faster than semantically unrelated words after mediated retrieval, resulting in quicker reaction times on semantically related words than on words that were semantically unrelated to the keyword mediator. In contrast, after direct retrieval such an advantage would not occur because the keyword had not been activated prior to the lexical decision task. Therefore, the semantic priming effect could be used to study changes in the activation of keyword mediators.

The study consisted of several experiments. In a first vocabulary learning experiment, in which priming was measured shortly after initial encoding with keywords, Kole and Healy (2013) indeed found a mediated priming effect, indicating that the retrieval of the word meaning was initially mediated. In a follow-up experiment, the amount of repetitions of the words during practice was manipulated. There were three conditions in which participants were tested, namely after 5, 10 or 45 repetitions. The priming task indicated that the learners still used the keywords to retrieve the meaning of the target words after 5 repetitions. However, the keyword use had decreased dramatically after 10 and after 45 repetitions, as indicated by the absence of the mediated priming effect. The keyword use was similar after both 10 and 45 repetitions but differed significantly from the keyword use after 5 repetitions. Based on these findings the authors concluded that the learners still used the keywords when retrieving the correct translation after repeating newly learned words 5 times. However, after 10 repetitions, the retrieval of the words seemed direct instead of mediated, since there was no mediated priming effect anymore. These findings provide evidence in favour of a shift from mediated retrieval to direct retrieval. It is concluded that extended practice strengthens the connection between the foreign word and its translation in a way that the association with the keyword eventually disappears and a direct link arises. This shift to direct retrieval supposedly reflects a change in the word representations in memory and can be explained by means of the identical elements model. This model states that the initially mediated retrieval mainly strengthens the connections around the mediator, namely between the target form and the mediator, and between the mediator and the target meaning. However, in case of direct retrieval the target meaning is retrieved without activating the mediator, hence direct retrieval mainly strengthens the connection between the target form and its meaning.

In summary, these previous studies suggest that a shift from mediated to direct retrieval of the word meaning appears after multiple repetitions. A shift to direct retrieval has been argued to be a crucial part of successful learning in all skill acquisition (Rickard, 1997). Therefore, rather than models that suppose a change only in the strength of the representation (e.g., holistic strengthening models; covert mediation model), models that propose a change in the fundamental structure of the memory representation over the course of practice (e.g., direct model; IE model) seem to be more plausible in order to explain the underlying memory structure after repeated retrieval practice. Therefore, these direct models will be used as a starting point in the present study.

Think aloud method

The main focus in the literature on the keyword method has been to investigate this learning strategy by accuracy measures and reaction times, both during practice and on performance tests (e.g., Atkinson, 1975; Fritz et al., 2007; Kole, & Healy, 2013). However, previous studies did not focus on the exact thinking processes of the learners as they arise during the retrieval of words. In order to gain insight in the exact thinking processes of learners when learning vocabulary using the keyword method, and to further explore the presence of a possible shift from initially mediated to later direct retrieval, it is necessary to collect data in a different way. This alternative method will now be thoroughly discussed.

General description

Verbal reports can be used to gain insight in the thinking process of learners. In particular the think aloud method can be a very useful way to investigate these thinking processes. This method seems to be especially efficient when the thoughts are already present in the mind in verbal form, as is the case when learning and rehearsing words. When thinking aloud, learners verbalize their thoughts as they arise, directly from the conscious, short-term memory

(Ericsson, & Simon, 1993). Therefore, this method can only be used to study the thoughts that are in the awareness of learners. As for the present study the focus will be on these conscious thoughts, in attempt to reveal the cognitive processes that underlie vocabulary acquisition by means of the keyword method.

The think aloud method (Someren, Barnard, & Sandberg, 1994) is a frequently and broadly used research method. However, there have been some comments on the usefulness of this method. One point of criticism is the completeness of the observation of the thinking process, since only the conscious thinking process and not the unconscious thinking process can be verbalized. Another point of criticism is that verbalizing every thought might slow down the thinking process, since the working memory has to perform two tasks at once. Therefore it has been argued that thinking aloud might not provide an accurate reflection of the natural thinking process (Jääskeläinen, 2010).

These and other concerns have been thoroughly discussed by Kumar (2017). First, the extent to which verbalizations describe the actual cognitive process is addressed. Even though the unconscious thinking process cannot be directly reported by means of the think aloud method, it does provide accurate information about the conscious thinking process, which can be used to explain the unconscious thinking process. Further, concerning the slowing down of the thinking process it is argued that the think aloud method could indeed slow down the response time of the learners. However, it will not fundamentally change the nature of the cognitive process and therefore it still gives an accurate reflection of the natural thinking process. Hence, it is concluded that these two points of criticism do not damage the validity of the think aloud method.

Additionally, in order to collect reliable think aloud data, Kumar (2017) argues that some important methodological aspects need to be taken into account. That is, the differences between participants need to be considered, since not all participants can verbalize their thoughts with the same ease. Also, warm-up sessions are necessary in order for the participants to get used to thinking aloud. Further, the choice of topic should be taken into account, since not every topic is suitable for the think aloud method. Lastly, the observer effect should be considered. That is, the presence of an experimenter might impede the natural course of the thinking process, since participants might feel extra aware and generate socially acceptable responses. In order to collect reliable data on the thinking process of participants, the present study will take all of these methodological aspects into account.

Limitations of previous think aloud studies

The think aloud method has already been used to investigate the keyword method in some previous studies, but perhaps not in the most ideal and reliable manner. For instance, Adams and McIntyre (1967) used verbal protocols to study mediator use and collected these verbal data after the retrieval task had finished. The participants could indicate the thoughts they had during the retrieval in a unilateral way by choosing from several options (e.g., whether they used a mediator, directly retrieved the target, guessed, or did not know how they retrieved the target word anymore). Even though these verbal reports did manage to gain insight in the mediator use by showing a decrease in keyword mediator use over the course of repetition, this method lacks reliability because of the unilateral reporting options and since the verbal reports were collected after the retrieval task.

Likewise, Crutcher and Ericsson (2000) used retrospective verbal reports in order to determine whether the participants used the keywords or not. The participants had to report their thoughts, not during, but immediately after they performed the retrieval task. So, they would first submit their answer on the retrieval task and later report the thoughts they had when they performed the retrieval task. However, the collection of the verbal reports immediately after the task might entail problems concerning the completeness of the verbal reports. Namely,

when people report their thoughts after they emerge, participants tend to report a personal view on the thinking process instead of the thoughts exactly as they emerged (Kumar, 2017). Collecting verbal reports not during, but after the task might have provided an incomplete and unreliable reflection of the thinking process. Thus, even though the retrospective reports of Crutcher and Ericsson (2000) were less unilateral than those of Adams and McIntyre (1967), the usefulness of these retrospective reports is questionable. Therefore, retrospection might not be a reliable way to investigate the use of mediators.

Alternatively, concurrent protocols could provide a more accurate image of the thinking process than the verbal protocols of Adams and McIntyre (1967) and the retrospective verbalizations of Crutcher and Ericsson (2000). Concurrent protocols, also known as thinking aloud, could provide a more complete image of the thinking process. In concurrent protocols participants express all the thoughts that come to mind during the retrieval task. The verbalizations will be more complete, since they are not collected after the memory task, but during the memory task itself. Other than response speed, which will not be investigated in the present study, the retrieval process will not be fundamentally hindered by this unnatural way of thinking (Kumar, 2017). Furthermore, concurrent verbalizations can give an accurate view of the thinking process, since it follows the stream of information in the brain rather precisely (Jääskeläinen, 2010; Kumar, 2017).

To our knowledge, concurrent protocols have not yet been used in studies on the keyword method. In many studies on the keyword method the verbal reports were elicited after the primary memory task had been performed (e.g., Adams, & McIntyre, 1967; Richardson, 1998; Crutcher, & Ericsson, 2000). Also, many studies elicited verbal reports on the specific recall strategies that had been used (e.g., Ellis, & Beaton, 1993b), instead of focussing on the exact thoughts during the recall process. By means of concurrent protocols the present study aims to tackle these shortcomings of previous studies and provide a valid reflection of the thinking process.

Present study

There seems to be general agreement about the positive effects of the keyword method (e.g., Atkinson, 1975; Crutcher, & Ericsson, 2000; Kole & Healy, 2013). Learning words while using the keyword method seems to result in a high retention of these words, especially when this method is combined with retrieval practice (Fritz et al., 2007). Prior research suggests that the activation of the keyword mediators changes when the representations of the word form and meaning are strengthened through practice. However, it is still unclear when such changes in mediator activation appear and whether these changes are beneficial for learning.

Several previous studies aimed to clarify the underlying structure of the memory representations of words that are learned using the keyword method. Most of these studies explain the underlying representation initially by means of a mediated model, in which the target meaning is recalled by activating the keyword mediator first. After a certain amount of practice however, a shift is assumed to appear from a mediated model to a direct model, which states that the keyword is not activated anymore, but the meaning of the target words can be retrieved directly (Rickard, & Bajic, 2003; Kole, & Healy, 2013). Despite evidence for a shift from mediated to direct retrieval, it is not clear yet when this shift exactly occurs. Most studies that investigated this shift focussed on the actual appearance of the shift rather than the exact moment of the shift (Crutcher, & Ericsson, 2000; Rickard, & Bajic, 2003; Kole, & Healy, 2013). Therefore, the moment of the shift from mediated to direct retrieval is still unclear and the first aim of the present study is to determine this exact moment of shift.

The second issue that is addressed in this study is whether the shift from initially mediated to direct retrieval is related to the retention of word knowledge over time. Despite the

evidence for the effectiveness of the keyword method (Atkinson, 1975; Crutcher, & Ericsson, 2000; Fritz et al., 2007; Kole & Healy, 2013), not everyone is convinced by its usability on a longer term. Wang, Thomas, and Ouellette (1992), for instance, performed a word learning study in which the words were initially encoded by means of keyword mediators, and then repeatedly practice in four practice rounds. The results showed that words that were learned using the keyword method were retained better after one week than words that were learned by means of restudying, but only in case there was a test immediately after the practice session. In case there was no immediate test, in other words if there was no round of retrieval practice, the restudying group outperformed the keyword group. Based on these findings the authors doubt the long-term effects of the keyword method, since they have been partially outperformed by the restudying group. However, the practice phase contained only four repetitions of the novel words. Possibly, the moment of shift from mediated to direct retrieval could influence the long-term retention of the words. Therefore, an alternative explanation for these low retentions might be that the moment of shift had not been reached yet and the word representations, which were still mediated, might have been weaker and less resistant to forgetting.

The present study aims to gain more insight in the effectiveness of the keyword method on the longer term. The retention of the words that were learned by means of keyword mediators might relate to the structure of the memory representation that has been established during practice and accordingly to the way in which the words are retrieved during practice, either with mediated or direct retrieval. Rickard and Bajic (2003) argue that word retrieval may be hindered by mediated retrieval in terms of recall speed and accuracy. Direct retrieval usually allows a quick and errorless retrieval of the word, whereas the extra step in mediated retrieval could result in a slower retrieval, containing more errors. This could have considerable impact on the language acquisition of the learners. For instance, the learners could have problems remembering which competitor was the correct one, since they can get confused by the keyword and other (semantically) related competitors. As Dunlosky et al. (2013) describe: “For instance, when a student retrieves “a dentist holding a large molar with a pair of pliers,” he or she may have difficulty deciding whether the target is “molar,” “tooth,” “pliers,” or “enamel.”” (p. 23). However, when learners shift to a direct retrieval this confusion most likely disappears, since there is no mediator that stands in between the target form and its meaning. Thus, in case the learners shift to direct retrieval, this confusion would disappear and it would not hinder the effectiveness of the keyword method.

The keyword method will be thoroughly investigated by means of the think aloud method. Dutch college students will be presented with novel Swahili vocabulary together with corresponding keywords and mnemonic sentences, in order to learn the meaning of the Swahili words. We chose to provide the mnemonics to the participants, because the keyword method is more effective when the participants are provided with the keywords and mnemonic associations instead of generating them on their own, provided that these are qualitatively good mediators (for an overview see Pressley, Levin, & Delaney, 1982). Also, when participants could choose whether they wanted to be provided with a keyword or not, it was found that the keywords were requested most (89%) of the time (Atkinson, 1975). So there seems to be an advantage of provided compared to self-generated keywords and mnemonic associations. An additional advantage of providing the mnemonics to the participants is that it ensures that the keyword and mnemonic use is consistent across all participants. Therefore, it strengthens the validity of the comparison of the use of the keyword method between the participants.

The present study will make use of the keyword method in combination with retrieval practice. This combination is argued to be a very effective strategy for foreign vocabulary learning (Fritz et al., 2007). The learners will practice the words, that they have learned by means of the keyword method, repeatedly while saying aloud every thought that comes to mind. Based on these verbalizations it will be determined whether the participants use a keyword or

not. In this way it is aimed to determine the exact moment of the shift from mediated to direct retrieval and to gain insight in the duration of the keyword use. Furthermore, investigating and determining the moment of shift might provide insight in the amount of repeated practice that is necessary in order to establish a direct link between the target form and its meaning. This would allow us to make recommendations about the use of the keyword method, especially about how many repetitions are needed in order to establish a thorough representation of the word in the memory.

The learners will be tested on their word retention after one week, which will allow us to investigate the effectiveness of the keyword method on a longer term. Subsequently, we aim to make predictions about the retention after one week based on the moment of the shift from mediated to direct retrieval on both participant level and word level. In other words, it is intended to predict the performance of the learners based on their shift behaviour and to predict the retention of individual words based on their individual corresponding shift moments. Investigating the retention of words after one week might provide evidence supporting the effectiveness of the keyword method on a longer term. The retention after one week might also provide information about the use and structure of the keyword method, based on which a recommendation about the necessary amount of repetitions for an effective recall of the foreign words on a longer term could be made.

Therefore, the present study aims to answer the following question: When does the shift from mediated to direct retrieval exactly occur when learning vocabulary using the keyword method, and can this shift moment predict the future retention of the words that were learned using the keyword method? This question can be divided into three sub-questions. The first sub-question addresses the precise moment of shift from mediated retrieval to direct retrieval. The second sub-question addresses whether the individual shift behaviour of learners could predict the future retention of the words for these specific learners. In other words, the question is whether learners who use the keywords for a longer time differ from learners who use the keywords for a shorter time in the amount of words that they remember after one week. The third question addresses the retention of individual words and investigates whether the shift moment of a word predicts the retention of this specific word later on. More specifically, are words that are retrieved using keywords for a longer time remembered better or worse than words that are retrieved using keywords for a shorter time? For reasons of clarification these three questions are summed up below.

1. When do learners shift from mediated to direct retrieval?
2. Do shift differences between learners predict differences in performance later on?
3. Do shift differences between words predict differences in performance later on?

The expectations concerning the first research question are based on the study of Kole and Healy (2013). They found that learners still used the keywords when retrieving the correct translations after repeating the novel words 5 times. However, after 10 repetitions they observed that the learners made use of direct retrieval instead of mediated retrieval. Based on these findings it is expected that the average moment of shift from mediated to direct retrieval should be somewhere in between 5 and 10 repetitions.

Regarding the second and third research question, it is expected that an early shift from mediated retrieval to direct retrieval leads to a more accurate retention of the words later on. Learners with an early shift will supposedly remember more words than learners who shift at a later moment. Accordingly, words that have an early shift will be retrieved more accurately than words with a later shift. These expectations are based on the idea the retrieval of the words shifts from initially mediated retrieval to direct retrieval, which is established after repeated practice (Rickard, & Bajic, 2003; Kole, & Healy, 2013) and that this shift is crucial for learning

(Rickard, 1997). Consequently, a direct mapping of the target form onto the meaning of the target occurs. This idea of direct retrieval is adopted in direct models on memory retrieval, like the identical elements model (Rickard, & Bajic, 2006), in which it is proposed that repeated practice leads to the establishment of individual memory representations, containing only the practiced components. As for the present study, it is expected that independent representations will be established containing only the foreign word and its translation. This direct retrieval is said to be most viable since the retrieval process consists of only one step, in which the relevant information can be directly retrieved from memory, whereas mediated retrieval requires extra steps in which the relevant information is retrieved through other, possibly irrelevant and confusing information which could hinder the retrieval of the correct meaning (McElree, 2001; Dunlosky et al., 2013).

Method

Participants

In total, 33 participants were tested. All participants were college students and native speakers of Dutch without any previous knowledge of Swahili. The data of 3 participants had to be removed from the analyses, because these participants did not perform the task as was intended. These exclusions will be thoroughly discussed in the results section. The remaining 21 female and 9 male participants in the present study had an average age of 22.6 years old ($SD = 2.1$) and ranged between 18 and 28 years old.

Overview of the experiment

The present study consisted of two sessions. The first session contained a practice phase with 2 initial encoding blocks and 10 retrieval blocks. The second session took place after a 6 to 8 day delay ($M = 6.8$, $SD = 0.8$) and was a test session in which the participants were tested on both their receptive and productive vocabulary knowledge.

Design

The dependent variable in the present study was the performance on the test one week after practice (three measures: perceptive performance, leniently scored productive performance, strictly scored productive performance). There were three independent variables in the present study. The first independent variable was retrieval type, which could either be mediated or direct: *mediated retrieval* in case a word was translated by using the keywords with which they had learned the novel words, and *direct retrieval* in case a word was translated without using the keyword. The second independent variable was the moment of shift from mediated to direct retrieval, and the third independent variable was the condition of the words, which could either be the experimental condition or the control condition.

Materials

Stimuli Selection. In the present study 50 Swahili words were used, with for every word a corresponding Dutch translation, an orthographically similar keyword, and a mnemonic association sentence (see Table 1 for a complete overview of the stimuli). The stimuli that were used in the practice session were the same as the stimuli in the test session of the experiment.

Of these 50 stimuli, 34 were retrieved from an already existing dataset. In a previous study participants had to fill in mnemonics, which resulted in a dataset with mnemonics that were generated by the participants themselves (van den Broek, Takashima, Segers, & Verhoeven, in press). Within this dataset we made a selection of suitable stimuli. The first selection criterion was that there had to be a consistency between the mnemonics across participants, in a way that a Swahili word was selected when there were various participants who used the same corresponding keyword. Other criteria were that these Swahili words had to be pronounceable in Dutch, so that it was possible to create a corresponding Dutch keyword, and that they were imaginable, so that they could be used to create a mental image. The other 16 Swahili words were selected from another dataset (Nelson & Dunlosky, 1994), again based on the criteria that the words had to be pronounceable and imaginable.

Pilot. Several small pilots were performed in order to ensure the quality of the stimuli. In the first pilot, 10 participants generated keywords corresponding with the selected Swahili words, without knowing the meaning of these words. They were instructed to write down keywords based on orthographic and phonological similarity. Their responses were compared to the list of keywords that was initially created for the present study. It appeared that most of the keywords in the pilot were exactly the same as the ones in our stimuli list. Only a few of the initial Swahili words in the list were adjusted because there was a strong consistency among participants about another keyword. However, most of the keywords that were mentioned in our pilot were already present in our stimuli list. This consistency between the keywords of these 10 additional participants and the initial keywords provided support for the effectiveness of our stimuli.

In the second pilot, the mnemonic association sentences were addressed. In this pilot, 6 participants were asked to rate the association sentences on their effectiveness on a scale of 1 (very bad) to 5 (very good). Based on these ratings, the sentences with a low average rating were adjusted. In general, the ratings of the sentences were good, with an above average rating ($M = 3.8$, $SD = 0.5$). However, three sentences had a low average rating ($M < 3$) and therefore had to be replaced with another sentence.

Lastly, in order to test the final experiment, one pilot participant performed the entire experiment. Based on the responses and comments of this pilot participants, some small additional adjustments were made to the stimuli and to some elements of the experiment, like minor adjustments in the instructions.

Criteria Keywords. Several criteria were used for the development of the keywords. First, the keywords needed to have substantial orthographic or phonological overlap with the Swahili words (e.g., *kofia* (hat) overlapped with the Dutch keyword *koffie* (coffee)). The second criterion was that the Dutch keywords had to be well-known words that were familiar to the participants. Support for the familiarity of the stimuli came from our pilot study, because all words were familiar to the pilot participants. Another applied criterion for the keywords was concreteness. The keywords in the present study were all concrete nouns, since the keyword method is most effective for concrete nouns and less effective for abstract nouns (Atkinson, 1975; Paivio, 1969; Hulstijn, 1997). However, not all target words were concrete. Some of the target words were abstract nouns, but still imaginable to some extent (e.g., *holiday* and *heaven*). In addition, both the imaginability of the keyword and the imaginability of an interaction between the keyword and the target meaning was taken into account. In order to ensure the effectiveness of the keyword method it is important that the mental images that correspond with specific keyword-target combination are vividly imaginable (Atkinson, 1975). Therefore, the keywords were selected only in case it was possible to form a vivid and clear mental image.

Another criterion was to ensure that both the keywords and translations were distinctly different from each other, so that the keywords would not be confusing. Initially both the Swahili words *bamba* (plate) and *pamba* (wool) were part of the stimuli set. However, our pilot

showed that these words were too similar and often caused confusion. Therefore, it was decided that only *pamba* stayed in the stimuli set and that *bamba* was excluded from the stimuli. Not only the Swahili words, but also the translations needed to be distinctly different. For instance, the initial stimuli set contained both *barua* (letter) and *ramani* (card). Our pilot showed that these were often confused as well, since the meaning of these two translations were rather similar. In order to resolve this matter of confusion, the meaning *card* was further specified to *map*, making these two translations distinctly different from one another.

Furthermore, there were some keyword in the stimuli set that contained names of a well-known brand (e.g., IKEA) or celebrity (e.g., Lady Gaga). Our pilot on the stimuli words showed that these words were known by all pilot participants, hence the assumption was made that the participants would probably know these famous and broadly used names. It is important to bear in mind that in case there were still some keywords that were unknown or unusable for specific participants, they could indicate this in the questionnaire at the end of the experiment and consequently the data of these words for these specific participants were excluded.

Criteria Mnemonic Association Sentences. Just like the keywords, also the mnemonic association sentences had to meet certain criteria. The first criterium for the sentences was that it had to be possible to combine the keyword and the translation in a vividly imaginable interaction. For instance, the association sentence that is suitable for the Swahili word *nyanya* (tomato) and Dutch keyword *ninja* (ninja) is *De ninja hakt de tomaat doormidden* (The ninja chops the tomato in half). This sentence forms a vivid interaction of the keyword and the translation and is likely to be imaginable for the participants. Another criterium was that the sentences had to contain only two nouns. Besides the keyword and the translation, it is preferable that there were no other nouns in the sentence, since this could be confusing for the participants and could lead to accidentally mixing up the wrong nouns. However, for 3 sentences in the stimuli set an additional noun was necessary in order for the sentence to be clear and semantically correct. Additionally, several other characteristics of the sentences were taken into account: all sentences were kept as short as possible so that there would be not additional and irrelevant information in the sentences (the length of the sentences ranged from 4 to 9 words, $M = 7.76$, $SD = 1.03$), all sentences had to be similar in complexity and therefore the dataset only contained active and no passive sentences, all sentences were written in the present tense, and lastly the position of the keyword and the translation were consistent in all sentences, in a way that the keyword was always placed before the translation.

Table 1

Overview of all stimuli: 50 Swahili words with the corresponding translations, keywords and mnemonic association sentences.

	Swahili	Translation	Keyword	Mnemonic association sentence
1	anga	<i>hemel</i> (heaven)	<i>engel</i> (angel)	<i>Ik zie een engel naar de hemel zweven</i> (I see an angel floating to heaven)
2	askari	<i>politie</i> (police)	<i>asbak</i> (ashtray)	<i>Hij gooit een asbak naar de politie</i> (He throws an ashtray at the police)
3	bandari	<i>haven</i> (harbour)	<i>bandana</i> (bandana)	<i>Ik doe mijn bandana om in de haven</i> (I put on my bandana in the harbor)

4	barua	<i>brief</i> (letter)	<i>bar</i> (bar)	<i>Hij schrijft aan de bar een brief</i> (He writes a letter in the bar)
5	basila	<i>bus</i> (bus)	<i>basilicum</i> (basil)	<i>Het ligt helemaal vol met basilicum in de bus</i> (The bus is full of basil)
6	chanjo	<i>schaar</i> (scissors)	<i>banjo</i> (banjo)	<i>Ik maak mijn banjo kapot met een schaar</i> (I destroy my banjo with a pair of scissors)
7	degaga	<i>bril</i> (glasses)	<i>Lady gaga</i> (Lady gaga)	<i>Lady gaga draagt een grote roze bril</i> (Lady gaga is wearing big pink glasses)
8	duka	<i>winkel</i> (store)	<i>doek</i> (cloth)	<i>Ik koop het perfecte doek in de winkel</i> (I buy the perfect cloth in the store)
9	dunia	<i>aarde</i> (earth)	<i>duinen</i> (dunes)	<i>Ik vind de duinen de mooiste plek op aarde</i> (I think the dunes are the best place on earth)
10	embe	<i>mango</i> (mango)	<i>gember</i> (ginger)	<i>Ik strooi gember over mijn mango</i> (I sprinkle ginger on my mango)
11	farasi	<i>paard</i> (horse)	<i>Ferrari</i> (Ferrari)	<i>De Ferrari rijdt hard achter het paard aan</i> (The Ferrari drives quickly behind the horse)
12	flava	<i>muziek</i> (music)	<i>lava</i> (lava)	<i>Het sissen van de lava klinkt als muziek</i> (The hissing of the lava sounds like music)
13	funguo	<i>sleutel</i> (key)	<i>funghi</i> (funghi)	<i>Op mijn pizza funghi ligt de verloren sleutel</i> (The lost key is on my pizza funghi)
14	godoro	<i>matras</i> (mattress)	<i>god</i> (god)	<i>Ik zie een god op het matras liggen</i> (I see a god on the mattress)
15	goti	<i>knie</i> (knee)	<i>goot</i> (gutter)	<i>Ik val hard in de goot met mijn knie</i> (I am falling in the gutter with my knee)
16	harafu	<i>geur</i> (smell)	<i>haar</i> (hair)	<i>Zijn haar verspreidt een hele vieze geur</i> (His hair spreads a very bad smell)
17	hewa	<i>lucht</i> (sky)	<i>HEMA</i> (name of a common Dutch store)	<i>Ik zie de HEMA door de lucht zweven</i> (I see the HEMA floating in the air)
18	jibini	<i>kaas</i> (cheese)	<i>bikini</i> (bikini)	<i>Ik stop mijn bikini helemaal vol met kaas</i> (I completely stuff my bikini with cheese)
19	kamba	<i>touw</i> (rope)	<i>kam</i> (comb)	<i>Ik hang mijn kam aan een lang touw</i> (I hang down my comb on a long rope)
20	kanisa	<i>kerk</i> (church)	<i>kan</i> (jug)	<i>Hij laat de kan vallen midden in de kerk</i> (He drops the jug in the middle of the church)
21	kaputula	<i>broek</i> (pants)	<i>katapult</i> (catapult)	<i>Met een katapult schiet ik mijn broek ver weg</i> (I shoot my pants far away with a catapult)
22	keja	<i>huis</i> (house)	<i>IKEA</i> (IKEA)	<i>Ik kan de IKEA zien vanuit mijn huis</i> (I can see the IKEA from my house)
23	kidonda	<i>wond</i> (wound)	<i>donder</i> (thunder)	<i>Door de donder barst mijn wond open</i> (The thunder bursts open my wound)
24	kioo	<i>spiegel</i> (mirror)	<i>kiosk</i> (kiosk)	<i>In de kiosk loop ik tegen een spiegel aan</i> (I walk against a mirror in the kiok)

25	kofia	hoed (hat)	koffie (coffee)	<i>Er stroomt koffie uit de hoed die ik opzet</i> (coffee is pouring out of the hat that I put on)
26	kujitia	sieraden (jewellery)	kuit (calf)	<i>Ik versier mijn kuit met sieraden</i> (I decorate my calf with jewellery)
27	leso	sjaal (scarf)	les (class)	<i>In de les draag ik altijd een sjaal</i> (In class I always wear a scarf)
28	likizo	vakantie (holiday)	likeur (liqueur)	<i>Ik drink heel veel likeur op vakantie</i> (I drink a lot of liqueur on holiday)
29	maiti	lichaam (body)	mais (corn)	<i>Ik gebruik mais om mijn lichaam te bedekken</i> (I use corn to cover my body)
30	mashua	boot (boat)	mast (mast)	<i>Ik klim helemaal bovenin de mast van de boot</i> (I climb completely in the top of the mast of the boat)
31	maziwa	melk (milk)	Maxima (name of the Dutch queen)	<i>Koningin Maxima drinkt melk</i> (Queen Maxima drinks milk)
32	meza	tafel (table)	mes (knife)	<i>Het mes blijft rechtop staan in de tafel</i> (The knife stays upright on the table)
33	mit	boom (tree)	mist (fog)	<i>Door de mist zie ik de boom niet meer</i> (I can't see the tree anymore through the fog)
34	mkate	brood (bread)	kater (tomcat)	<i>Mijn stomme kater gaat altijd op het brood zitten</i> (My stupid tomcat always sits on the bread)
35	moto	vuur (fire)	motor (motor)	<i>Mijn motor spuwt vuur als ik ermee rijd</i> (My motor spits fire when I drive it)
36	ndege	vliegtuig (airplane)	deeg (dough)	<i>Ik neem het deeg mee in het vliegtuig</i> (I take the dough with me on the airplane)
37	nyanya	tomaat (tomato)	ninja (ninja)	<i>De ninja hakt de tomaat doormidden</i> (The ninja chops the tomato in half)
38	paka	kat (cat)	pak (suit)	<i>Dit pak draagt mijn kat als het koud is</i> (My cat wears this suit when it is cold)
39	pamba	wol (wool)	panda (panda)	<i>De schattige panda slaapt tussen de wol</i> (The cute panda sleeps between the wool)
40	panya	muis (mouse)	ranja (syrup)	<i>De fles ranja valt bijna op de muis</i> (The bottle of syrup almost falls upon the mouse)
41	pombe	bier (beer)	pomp (pump)	<i>Uit de pomp stroomt warm bier</i> (Warm beer is pouring out of the pump)
42	ramani	landkaart (map)	raam (window)	<i>Midden op het raam tekent hij de kaart</i> (He draws the map in the middle of the window)
43	saduku	doos (box)	sudoku (sudoku)	<i>Er zitten heel veel sudokus in deze doos</i> (There are a lot of sudokus in this box)
44	simu	telefoon (telephone)	simkaart (simcard)	<i>Ik stop mijn simkaart in de telefoon</i> (I put my simcard in the telephone)
45	skati	rok (skirt)	skater (skater)	<i>De skater draagt thuis altijd een rok</i> (The skater always wears a skirt at home)

46	taa	<i>lamp</i> (lamp)	<i>taart</i> (cake)	<i>De taart is versierd met een lamp</i> (The cake is decorated with a lamp)
47	theluji	<i>sneeuw</i> (snow)	<i>thee</i> (tea)	<i>Ik laat mijn thee vallen in de sneeuw</i> (I drop my tea in the snow)
48	tofaa	<i>appel</i> (apple)	<i>toffee</i> (toffee)	<i>Ik doe gesmolten toffee over my apple</i> (I put melted toffee on my apple)
49	tunda	<i>fruit</i> (fruit)	<i>toendra</i> (tundra)	<i>Ik ga naar de toendra om fruit te plukken</i> (I go to the tundra to pick fruit)
50	zulia	<i>tapijt</i> (carpet)	<i>zuil</i> (pillar)	<i>De grote zuil staat op het tapijt</i> (The big pillar is on the carpet)

Procedure

The experiment took place in the Centre for Language Studies lab of the Radboud University in Nijmegen. The participants were tested individually in a small experiment booth. The present study consisted of two sessions, the first session contained the practice phase and took approximately 1 hour and 30 minutes. The practice phase included an encoding phase, containing an initial encoding block and a rating block, and a retrieval phase consisting of 10 retrieval blocks with feedback. The second session contained the test phase and took approximately 20 minutes. The test phase contained one receptive test block and one productive test block, both without feedback. The second session took place about one week (6 to 8 days) after the first session.

Practice Session. Initially, the participants were asked to fill out two informed consent forms, in which they gave permission to participate in the study, permission that their data could be used for scientific purposes, and permission to record their speech during some parts of the experiment. In these forms the anonymity and confidentiality of the personal information and responses of the participants were emphasized and reassured (see Appendix).

Encoding phase. The first two blocks of the practice session were the initial encoding block and the rating block, in which it was intended to assure initial encoding of the Swahili words by means of the keyword method.

Initial Encoding Block. The first session started with some general instructions on the computer screen in which the keyword method and the importance of the visualization of a mental image was explained. Before the initial encoding block started, the participants were instructed to form a vivid visualization of each mnemonic association by means of the following written instructions: “Try to remember the word by forming an image in your mind of the mnemonics that is as vivid as possible. A vivid visualization helps to learn the words and to remember them more quickly”. After the instructions, the participants were presented with an example and subsequently the encoding phase started. In the initial encoding block the participants had to learn the words in combination with the corresponding keywords and mnemonic sentences. To control for possible effects of presentation order, the stimuli were presented to each participant in a different random order. The duration of the presentation of each individual trial in the encoding phase was fixed, so the participants were offered a limited amount of time to study the combinations. Halfway through the encoding phase, the participants had the possibility to take a short break.

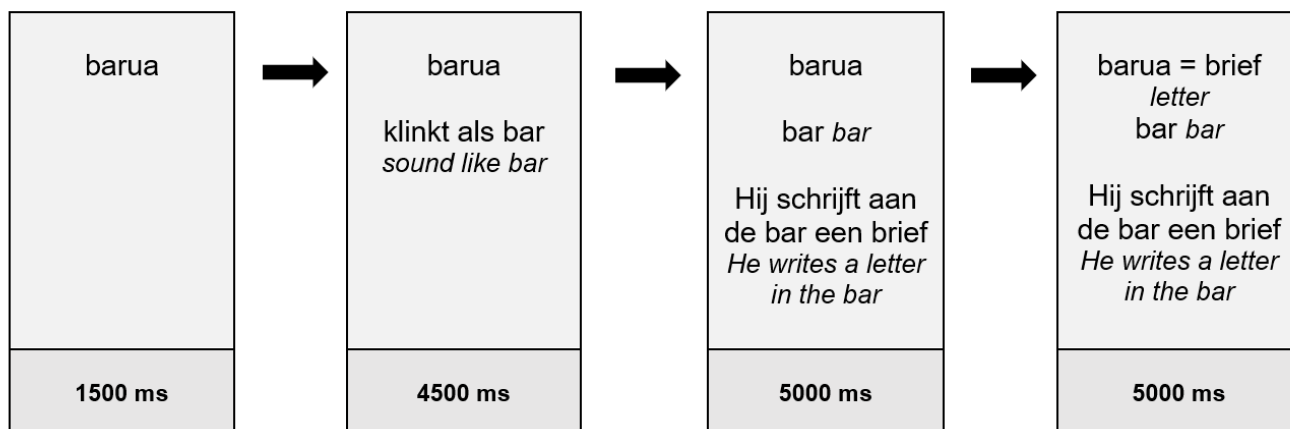


Figure 2. Presentation of the stimuli in the encoding phase, including their duration, with a total duration of 16000ms for each stimulus

All stimuli were presented to the participants on the computer screen. The Swahili word was always the first item that appeared on the screen and the participants had 1500ms to study this word before the keyword was added. The combination of the Swahili word and the keyword remained on the screen for another 4500ms. Subsequently, the mnemonic association sentence was added to the screen and this combination was presented for 5000ms. Lastly, also the translation of the Swahili word was added and this final combination was presented for another 5000ms. The total presentation duration of each stimulus was 16000ms, which allowed the participants to carefully study the words and imagine the associations that were described in the mnemonics. For reasons of clarification, the steps of the presentation in the encoding phase, including the presentation durations, are visualized in Figure 2.

Rating Block. After the initial encoding block, the participants proceeded to the rating block in which all the mnemonic association sentences were presented again in a different random order for each participant. The participants had to rate the imaginability of these associations on a rating scale. In other words, the participants were asked whether they had a clear image in mind for each mnemonic association sentence. The participants indicated their ratings by clicking with the mouse on a continuous scale from (1) *No, no image at all* to (4) *Yes, clear image*. The total duration of this rating block was approximately 10 minutes. Based on these ratings, each stimuli word was assigned to either the experimental or the control condition in such a way that the average of the ratings in the experimental condition ($M = 3.08$, $SD = 0.70$) and the control condition ($M = 3.08$, $SD = 0.71$) was similar. Because the participants were presented with all the stimuli again, this rating block was also a second encoding block for the participants.

Retrieval Blocks. In the retrieval blocks all words were retrieved and translated multiple times. This phase consisted of 10 retrieval blocks and took about 1 hour. The participants performed a receptive translation task in which the Swahili words were presented on the screen until the participants submitted a response by typing in the translation. After they submitted their response the participants immediately received feedback on their performance. In case the answer was correct, the submitted response turned green and an exclamation point was added to the response. However, if the response was incorrect, it turned red and the correct response was shown. In case of an incorrect answer, the participants were provided with both the mnemonic association sentence and the correct response in the initial two practice blocks. In all of the following blocks, only the correct response was provided, without the corresponding mnemonic association sentence.

Experimental and Control Condition. Of the 50 words, 30 were assigned to the experimental condition, and 20 to the control condition, with a different distribution of the stimuli words over the conditions for each participant. In this distribution the imaginability of the mnemonics was taken into account by means of the individual ratings of the participants. The items in the experimental condition were presented in all 10 retrieval blocks, and the items in the control condition were presented only in retrieval blocks 1, 2, 5, and 10. By presenting all the words of both conditions in the two initial blocks it was aimed to ensure sufficient initial encoding. After this initial encoding, the words in the control condition were only retrieved halfway through and in the end of the practice session, in order to make comparisons possible between the words after the same amount of time but after a different amount of retrieval practice since the initial blocks. In this way the effects of retrieval, rather than the effects of time, could be explored.

Think aloud Data. The think aloud method was used in order to obtain data about the keyword use of the participants in the retrieval blocks. This method required the participants to say aloud everything that crossed their mind while they were performing the translation task. Before the retrieval blocks started the participants received the following instructions on thinking aloud: “In the tasks in which you will think aloud, we ask you to say aloud every thought that comes to mind. The more complete you report your thoughts, the better. Concentrate on the task and try not to get distracted by thinking aloud. Please note that you should speak loud enough and do not keep your hand before your mouth while talking. If you still have questions about thinking aloud, you can ask them to the experimenter now”. In addition to these instructions on the screen, the experimenter gave a short explanation of the think aloud method in the general instruction prior to the experiment. The experimenter gave the exact following instruction to each participant: “When thinking aloud you have to say everything that comes to mind, no more and no less”.

After the instructions there was a short think aloud practice, in which participants had to calculate a sum while saying their thoughts out loud. The verbal responses of the participants were recorded with a microphone that was located next to the computer screen. During this practice exercise the experimenter sat in another room and listened through headphones to the responses on this practice exercise. The participants knew that the experimenter was listening during this think aloud exercise. The experimenter checked whether the participants had understood the task and performed the think aloud method in an accurate way, which was the case for all participants.

The retrieval block started immediately after the instructions and the exercise on the think aloud method. In order to ensure that the participants kept on thinking out loud, a written reminder was presented regularly during practice. In the initial two blocks, the reminder was presented after every third trial and in the rest of the practice blocks the reminder was presented regularly. The experimenter listened to the responses of the participant online during the task and kept track of the keyword use of the participants. The participants were not informed about the fact that the experimenter listened to their responses while they were performing the task, in order for the learners to feel more comfortable during the task and to prevent that the participants would give socially desirable responses (Kumar, 2017). In case participants forgot to think aloud for several trials in a row, the experimenter prompted the participants to think aloud by intercom, however these additional prompts were almost never necessary.

Test session. Approximately one week after the practice session, with a range between 6 and 8 days, the participants returned to the lab for the second session of the experiment. The second session only consisted of testing blocks. There were two tests, namely a receptive test and a productive test. The participants immediately started with the receptive test, in which the Swahili words were presented on the screen and they had to type in the Dutch translation while thinking aloud, just like in the practice phase of the first session. After the receptive test the

participants performed a non-linguistic distractor task, namely a digit span working memory task (Woods et al., 2011). In this task the participants heard a sequence of digits and then had to report this sequence by typing in the digits that they had just heard. This task was performed in order to reduce possible effects of the first, receptive test on the second, productive test. Subsequently the productive test started, in which the Dutch words were presented and the participants were asked to provide the Swahili translations while thinking aloud. The participants had not practiced the retrieval of the words this way around.

Data preparation

Scoring think aloud data. The verbal responses of the participants during the practice blocks were used in order to determine the moment of shift from mediated to direct retrieval. Each trial was scored on the following features: keyword use (whether the participant mentioned the keyword on each specific trial), keyword accuracy (whether the used keyword was the correct keyword), and mnemonic use (whether the participant mentioned the mnemonic association sentence). Besides these main scoring features, several other factors were kept track of in order to make additional observations and explanations about the responses of the participants during practice (e.g., the errors that were made; whether participants referred to the Swahili word; additional notes). Table 2 displays the scoring features on which the verbal responses of each trial were scored by means of an example of the acquisition of the Swahili word *skati* (skirt), over the course of the practice blocks.

All trials were scored on the same features and by the same rater. After scoring the complete dataset ($n = 1500$), the same rater scored a subset of all trials ($n = 300$) again in order to substantiate the reliability of the scores. This subset contained 10 random trials of each participant and was scored on exactly the same features as the complete dataset. There was complete overlap between the scoring on the whole dataset and the additionally scored subset, indicating high test-retest reliability of the scores.

Shift from Mediated to Direct Retrieval during Practice. First, the moment on which participants shifted from mediated to direct retrieval was determined on word level. Based on the scorings on each trial (see Table 2 for an example), the keyword and mnemonic use were converted into classifications of mediated or direct retrieval. In case the participants referred to the keyword and/or mnemonic association sentence, the word was encoded as *mediated*, and in case the participant did not mention the keyword and/or mnemonic association sentence, the word was encoded as *direct*. The scores for all the trials per word were then taken together and based on these scores the moment of shift for each individual word was determined (e.g., the word in Table 2 would be scored as a shift moment of 5, since from block 5 the participant did not refer to the keyword and/or mnemonic association sentence anymore).

To illustrate, in case a word was retrieved in a mediated way at first, but this mediated retrieval was replaced by direct retrieval from a specific block until the end of the experiment, the block in which direct retrieval was used for the first time was designated as the shift block. For example, if a learner would use mediated retrieval (*M*) until block 4, and only direct retrieval (*D*) on that specific word from block 5 until the end of the experiment, the pattern of shift on this word would be *MMMMDDDDDD*, and the moment of shift on this specific word would be granted the number 5. In case the shift was not as straightforward, for instance if learners changed back and forth between mediated and direct retrieval, a word was classified only as shifted when the learner did not change back to mediated retrieval anymore until the end of the experiment. So, if a learner used mediated retrieval in until block 4, direct retrieval in block 5, mediated retrieval in block 6 and 7, and direct retrieval again in block 8, 9 and 10, the pattern of shift would be *MMMMDMMDDD*, and the shift moment was scored to occur in

Table 2

Example of the verbal responses and the corresponding scorings on the word 'skati' (skirt) for one specific participant throughout the practice blocks. The corresponding keyword was 'skater' (skater) and the corresponding mnemonic association sentence was 'De skater draagt thuis altijd een rok' (The skater always wears a skirt at home).

Block	Response	Keyword	Keyword correct	Mnemonic sentence	Swahili word	Errors	Notes
1	"skati skater een skater met een rok aan" <i>skati skater is wearing a skirt</i>	1	1	1	1	x	x
2	"skati skater met een rok aan" <i>skati skater wearing a skirt</i>	1	1	1	1	x	x
3	"skater met een rok aan" <i>skati skater wearing a skirt</i>	1	1	1	0	x	x
4	"skater met een rok aan" <i>skati skater wearing a skirt</i>	1	1	1	0	x	x
5	"skati rok" <i>skati skirt</i>	0	0	0	1	x	x
6	"skati rok" <i>skati skirt</i>	0	0	0	1	x	x
7	"rok" <i>skirt</i>	0	0	0	0	x	x
8	"rok" <i>skirt</i>	0	0	0	0	x	x
9	"skati rok" <i>skati skirt</i>	0	0	0	1	x	x
10	"skati rok" <i>skati skirt</i>	0	0	0	1	x	x

Note. The columns have the following meaning: keyword = whether a keyword was mentioned; keyword correct = accuracy of the produced keyword; mnemonic sentence = whether the mnemonic association sentence was mentioned; Swahili word = whether the Swahili word was mentioned; errors = the type of error for those words that were translated incorrectly; notes = additional remarks on the verbal responses

block 8. In case a word was translated incorrectly, that specific retrieval of the word was classified as if it had not shifted and therefore was scored as a mediated trial. Lastly, the average shift moment of each participant was calculated by averaging all of the shift moments for each word on which the participant shifted at some point during practice.

Performance on the Test. The accuracy of the test responses was scored using the Levenshtein Distance measure (Levenshtein, 1966). This measure describes how many letters need to be removed, added, or replaced to change one word into another word (e.g., **hause** to **houses** would have a Levenshtein Distance of 2). The accuracy of the responses on the receptive test of the experiment was based on a Levenshtein Distance of ≤ 2 , meaning that the difference between the target and the response was small, but that typing errors were scored as correct. The responses of the productive test were scored based on two different Levenshtein Distances. First, the productive test was leniently scored, based on a Levenshtein Distance of ≤ 2 , just like on the receptive test. Second, the productive test was strictly scored, based on a Levenshtein Distance of 0. This additional score provided a measure of performance based on completely correct responses, in other words in case of complete overlap between the Swahili target word and the response. In addition, the proportion of correct responses was calculated for each test.

In this way, the different exclusions, which will be discussed thoroughly in the results section, for each individual participant were taken into account in the performance scores.

Input for Analyses. Three different input files were constructed, corresponding with the three main questions of the present study. The first input file contained all the information concerning the moment of shift and the number of shifts of each participant. The second input file included the averaged information about the presence and the moment of the shift and the test performances on participant level, and the third input file contained the shift and performance information on word level. Together, these three files formed the input for the statistical analyses.

Analyses

Moment of Shift. In order to answer the first question concerning the moment of shift from mediated to direct retrieval, descriptive statistics were used to describe the observed moments of shift. The average moment of shift for the participants who shifted from mediated to direct retrieval at some point during the practice blocks was determined as follows. First, the moment of shift on all individual words was averaged for those words on which learners shifted, which indicated the average moment of shift for each individual participant. Next, these means were described using boxplots and descriptive statistics, including a group average to show after how many retrievals the total group of participants shifted from mediated to direct retrieval. Additionally, the development of the decrease in mediated retrieval over the course of the practice blocks was explored. The average proportion of mediated retrieval was calculated for each participant in every block. Next these proportions of mediated retrieval of all participants were averaged per block. In order to compare the proportions of mediated retrieval between the conditions in specific blocks, paired samples t-tests were performed.

Variations on Participant Level. Next, the question whether shift differences between learners could be a predictor of the performance on the later retention test was addressed. First, the proportion of correct responses on both the receptive and productive test were calculated for each participant. Next, the Pearson correlation was calculated between the average moment of shift and the performance on the final tests. In other words, the shift moment of those participants that did shift, was related to their test performance. A negative correlation would indicate that as the moment of shift decreases (i.e., the earlier the participant shifts), the performance on the later test increases. However, a positive correlation would indicate the opposite, namely that as the shift moment decreases, also the performance on the later test decreases. Besides, it could also be possible to find no correlation at all, indicating that there is no relation between the shift and the later retention.

Variations on Word level. Lastly, the question whether differences between words predict differences in later performance was addressed. First, the proportions of correct responses on the tests were calculated for both the mediated and the direct words in the experimental condition. On each test, the proportions of correct responses on the mediated words on were compared with the proportions on the direct words by means of paired samples t-tests. Further, the differences on word level were analysed by means of Logistic Mixed Effects Models. By means of the lme4 and plogis function in R (R Core Team, 2017) both the presence and the moment of shift on the words were related to the performance on the final tests on word level, taking the differences on both participants and learners into account. Specifically, it was aimed to determine whether shifted, directly retrieved words were remembered better than non-shifted, mediated words and whether words that shifted earlier were remembered better than words that shifted later, for those words on which a shift from mediated to direct retrieval did occur. In order to investigate this, the direct words were compared to the mediated words within the same participant, also the differences between the two conditions were explored. Thus, the

main predictors in this analysis were the retrieval type of the words, either direct or mediated, and the shift moment of the words. Further, in order to control for possible confounds both the learners and the words were included as random effects.

Results

The main questions in the present study concern the exact moment of shift from mediated to direct retrieval and the relation between this shift moment and the receptive and productive performance one week after the practice session, on both participant and word level. These questions are examined in three consecutive analyses, starting with the determination of the moment of shift, then the analysis on participant level and lastly the analysis on word level.

Exclusions

Prior to the analyses, the data of three participants had to be excluded. Two participants made use of a large number of mnemonics that they constructed themselves, namely at least one fifth (10 out of 50) of the mnemonics was adjusted or replaced completely. The third excluded participant performed the think aloud task in a manner that was not intended in the present experiment. Instead of saying aloud the thoughts on the words, this participant told several lengthy monologues that did not relate to the actual word learning task, in spite of prompts to report task-related thoughts. Thus, based on the lack of think aloud data about the word learning task, this participant had to be excluded as well.

Additionally, in the complete dataset of the 30 included participants, a total of 31 words were excluded from the analyses based on learner specific reasons. Most of these words were excluded because the participants had a strong association with a different keyword and replaced the provided keyword by another keyword (28 cases; for instance, several participants replaced the keyword *doek* (cloth) with the keyword *doekoe* (Dutch slang for money)). Further, some words were excluded because participants indicated that the Swahili word was already familiar (3 cases). This was either because they already knew the Swahili word (1 case) or because they knew the word from another language (2 cases; twice for the word *meza*, which has an equivalent in both Spanish and Slovenian). The average amount of excluded words for all remaining participants was 1.03 and ranged from a minimum of 0 to a maximum of 7 excluded words ($n_{pp} = 30$, $SD = 1.65$).

Manipulation check: imaginability rating of the mnemonics

The ratings of the participants were taken into account when assigning words to the experimental and control condition, in such a way that the average ratings in the two conditions were similar. However, an examination of the ratings of the individual participants on the mnemonics showed that the ratings did not necessarily reflect the differences in the imaginability of the items, but rather the differences in the attitude of the participants towards the items. The variation on the ratings between the participants ($SD = 0.29$) was higher than within the participants ($SD = 0.19$), indicating that the participants rated the mnemonics in dissimilar ways. Moreover, the Pearson correlation between the ratings and the performance on the receptive test on word level was not significant ($r = -0.03$, $p = 0.29$, $n_{words} = 1500$) and for the productive task on word level there was a significant correlation around zero ($r = -0.05$, $p = .05$, $n_{words} = 1500$), indicating that there is no correlation between the ratings and the productive performance. For these reasons, it was decided not to exclude words from the data

analysis based upon the ratings. That is, the ratings were not considered as a well-grounded instrument to base word exclusions on.

Manipulation check: accuracy during practice

The accuracy of the responses during practice was high. Calculated across all trials, participants responded with the correct translation on 96.5% of all the trials. In the experimental condition, 97.3% of the trials were answered correctly; in the control condition 93.4% of the trials were responded correctly. The proportion of correct responses changed over the course of practice from 84.2% in block 1, to 99.9% in block 10 in the experimental condition, and from 82.9% in block 1 to 97.8% in block 10 for the control condition.

Additionally, it was checked whether including only the correct trials, rather than both the correct and incorrect retrieval attempts, considerably affected the information about the shift behaviour of the participants. The complete dataset ($n_{words} = 1469$, $n_{pp} = 30$) was compared to a subset of the data, which contained only words that were answered correctly in all practice blocks ($n_{words} = 1193$, $n_{pp} = 30$). The possible variation between these two datasets was checked, because there might be a difference in the determination of the shift when focussing on only the correct retrievals and on both correct and incorrect retrieval attempts. For instance, when a word was scored as shifted after 5 retrieval attempts, but was not translated correctly during the first retrieval, a shift moment of 5 is accurate when focussing on the retrieval attempts, but the shift moment would be 4 when focussing merely on the correct retrievals.

A paired-samples t-test comparing the average shift moments of the participants over all words and the subset of correctly translated data showed a significant lower average shift moment for the correct subset ($M = 8.23$, $SD = 1.88$) than for the complete dataset ($M = 8.27$, $SD = 1.84$) in the experimental condition ($t(20) = 2.2192$, $p = .04$). No significant difference between the two datasets was found in the control condition ($t(16) = 1.1708$, $p = .26$). The present study chose to focus only on the retrieval attempts and therefore the complete dataset was used for the analyses. This decision provided a larger dataset for the analyses, with 276 additional words compared to the subset of only correctly retrieved words. In this way the sample size was enlarged which increased the power of the analyses. Moreover, the incorrect retrieval attempts did lead to additional practice of the correct translations by means of the provided feedback on the responses and should therefore be taken into account in the shift determination.

RQ1: When do learners shift from mediated to direct retrieval?

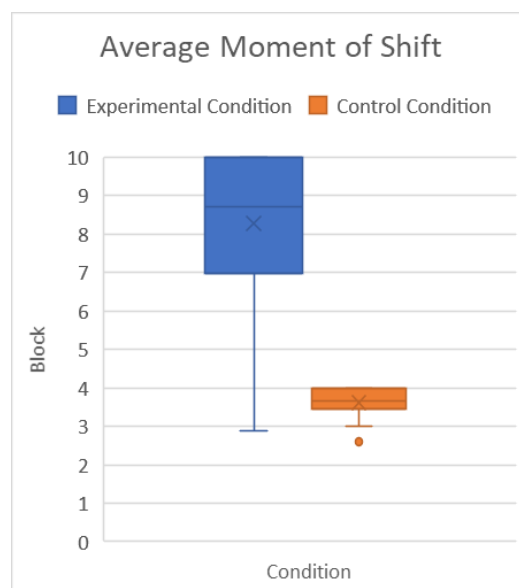
Proportion mediated and direct retrieval. The participants mostly used mediated retrieval in order to translate the words during the practice session. In the experimental condition, in which each word was retrieved 10 times, the participants retrieved on average 78.4% of the words from memory using mediated retrieval throughout the entire experiment ($SD = 26.7\%$, $n_{pp} = 30$). Similarly, in the control condition 85.9% of the words were translated using mediated retrieval ($SD = 18.0\%$, $n_{pp} = 30$). A paired-samples t-test showed that this difference in the amount of mediated retrieval between the conditions was significant ($t(29) = -2.6773$, $p = .01$).

Patterns of shifting. The information about the moment of shift was derived from the think aloud data (see method section). There were several possible shifting patterns. In case the words were scored as *mediated*, the pattern consisted of either mediated words (*M*) only (e.g., *MMMMMMMMMMMM*), or there were some direct retrievals (*D*) on a specific word that changed back to mediated retrieval in a later block and stayed mediated throughout the rest of the experiment (e.g., *MMMDMDMMMM*). In case the words were scored as *direct*, the words shifted from mediated to direct retrieval in a certain block and stayed directly retrieved

throughout the rest of the experiment (e.g., *MMMMDDDDDD*), or changed back and forth between mediated and direct retrieval but were directly retrieved from a certain block until the last block of the experiment (e.g., *MMMDMDMDDD*)¹. The word was shifted when the mediated retrieval changed to direct retrieval and stayed direct from this specific retrieval throughout the whole experiment. The moment of shift was determined by taking the specific block of retrieval in which the shift occurred.

Average moment of shift. In order to determine the appearance of the exact moment of shift from mediated to direct retrieval, only the data of the participants who did shift were included. In this subset of the participants that shifted for at least some part of the words, 69.2% of the words in the experimental condition ($SD = 27.1\%$, $n_{pp} = 21^2$) and 75.0% of the words in the control condition ($SD = 17.4\%$, $n_{pp} = 17^3$) were always translated using mediated retrieval during the practice blocks. These percentages were calculated on participant level and show the average proportion of mediated words for each individual participant by condition.

The distribution of the average moment of shift from mediated to direct retrieval on participant level in each condition is visualized in Figure 3 and further examination of the moment of shift is summarized in Table 3. In the experimental condition the participants shifted to direct retrieval on average after 8.27 retrievals ($SD = 1.84$, $n_{pp} = 21$). In the control condition the average shift to direct retrieval appeared after 3.5 retrievals ($SD = 0.40$, $n_{pp} = 17$). Besides information about the shift moments in the complete dataset, Table 3 contains two additional columns that present the shift information on the subset of the data that only contained words that were retrieved correctly throughout the entire practice phase.



Note. Means and SDs are presented in Table 3

Figure 3. Distribution of the average shift moment by condition on participant level for the shifted words.

¹ Note that this shift pattern is an example for a word in the experimental condition. Words in the control condition are retrieved only 4 times, so a possible shift pattern in the control condition would be *MMMM* for a mediated word, and *MMDD* for a direct word.

² 9 out of 30 participants did not shift to direct retrieval for any of the words in the experimental condition.

³ 13 out of 30 participants did not shift to direct retrieval for any of the words in the control condition.

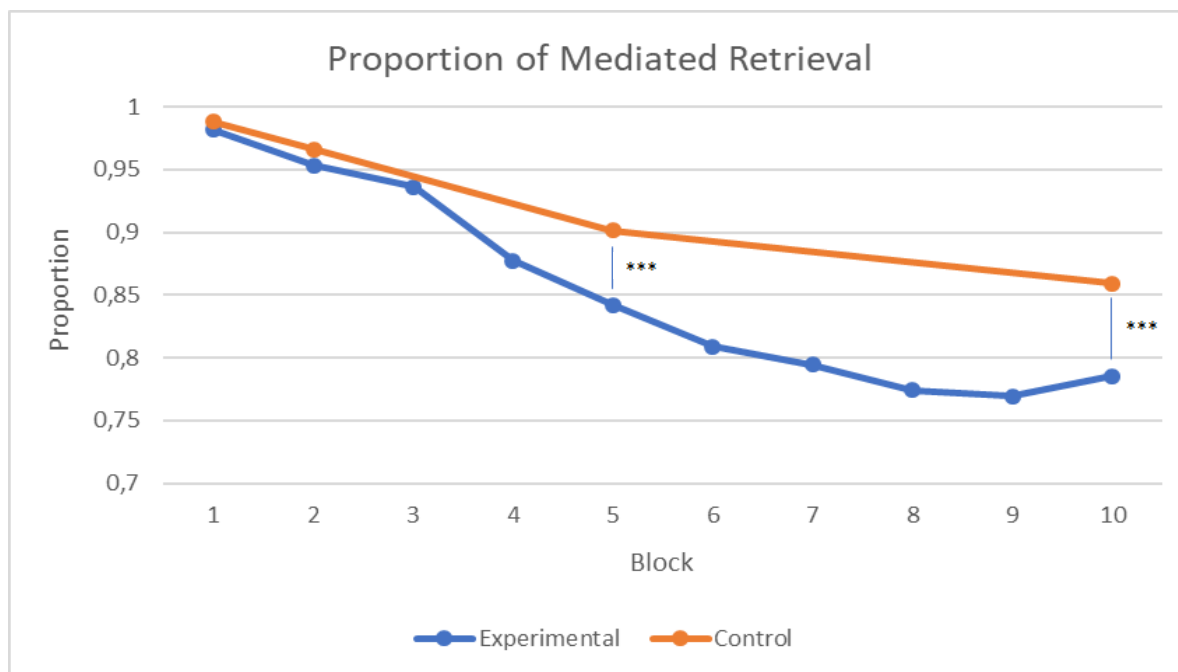
Table 3

Overview of the shift information: Average moment of shift, number of shifts and proportion of shifts on participant level by condition. The average moment of shift was calculated for both the complete data set and the dataset of merely correctly retrieved words, and the number and proportion of shift were only calculated for the complete dataset.

	Experiment Condition				Control Condition			
	Average Shift	Average Shift Correct Only	Number of Shifts	Proportion of Shifts	Average Shift	Average Shift Correct Only	Number of Shifts	Proportion of Shifts
	$n_{pp} = 21$	$n_{pp} = 21$	$n_{pp} = 21$	$n_{pp} = 21$	$n_{pp} = 17$	$n_{pp} = 17$	$n_{pp} = 17$	$n_{pp} = 17$
Minimal	2.89	2.79	1	0.03	2.60	2.00	1	0.05
1st Qu	7.14	7.00	3	0.10	3.50	3.40	3	0.15
Median	8.71	8.55	5	0.19	3.67	3.64	4	0.21
Mean (SD)	8.27 (1.84)	8.23 (1.88)	9 (7.81)	0.31 (0.27)	3.62 (0.40)	3.52 (0.59)	4.88 (3.48)	0.25 (0.17)
3rd Qu	10	10	14	0.47	4	4	5	0.25
Maximal	10	10	27	0.93	4	4	15	0.75

Note. Average Shift was calculated per participant, across words, and then aggregated for this table. Average Shift Correct Only was calculated in the same way, but including only those words that were retrieved correctly in all practice blocks. Number of Shifts was determined per participant as the number of words on which a shift appeared. Proportion of Shifts was calculated by dividing the number of shifts by the total number of trials, to allow comparison between the experimental and control condition, that differed in the total number of words they contained.

Proportion mediated retrieval per condition. The proportion of mediated retrieval decreased over the course of the practice blocks. Figure 4 displays the proportion of mediated retrieval in the practice phase of the experiment per block in each condition (see Table 4 for an overview of the observations in Figure 4). As Figure 4 indicates, a paired samples T-test confirmed that the proportion of mediated retrieval significantly decreased over the course of practice in both the experimental condition, $t(29) = -4.0054, p < .001$, and the control condition, $t(29) = -3.9302, p < .001$. Furthermore, the paired samples T-test showed that there was no difference in the proportion of mediated retrieval on the words between the conditions in the first practice block, $t(29) = 1.7255, p = .10$. However, there seemed to be a difference between the conditions in the proportion of mediated trials in the second block, $t(29) = 2.2759, p = .03$. This indicates that in the initial block the proportion of mediated retrieval develops similarly in the two conditions. However, starting from the second block the figure suggests that mediated retrieval decreased more strongly over the course of the practice blocks in the experimental condition (in which the words were retrieved 10 times) than in the control condition (in which the words were retrieved 4 times).



Note: significance level *** $p < .001$

Figure 4. Proportion of mediated retrieval by condition on word level over the course of practice.

Table 4

Total number of shifts per block in each condition and the total number of words by retrieval type in each condition.

Block	Appearance of Shift		Direct Retrieval		Mediated Retrieval	
	Experimental Condition	Control Condition	Experimental Condition	Control Condition	Experimental Condition	Control Condition
1	2	1	16	7	884	593
2	16	9	41	20	859	580
3	14	x	56	x	844	x
4	12	x	108	x	792	x
5	9	24	139	58	761	542
6	24	x	168	x	732	x
7	11	x	181	x	719	x
8	16	x	199	x	701	x
9	27	x	203	x	697	x
10	58	49	189	83	711	517
Total	189	83				

Note. The column Appearance of Shift shows the number of words that shifted from mediated to direct retrieval (i.e., were directly retrieved throughout the rest of the experiment) in each specific block. This column displays different words in each block. In contrast, the columns Direct Retrieval and Mediated Retrieval display the retrieval type (direct or mediated) of every word, regardless whether these words were shifted (i.e., directly retrieved throughout the rest of the experiment) or not. The values in the columns Direct Retrieval and Mediated Retrieval include cumulative scores.

Halfway through the experiment (block 5), the proportion of mediated retrieval was significantly smaller in the experimental condition than in the control condition, $t(29) = 3.7953$, $p < .001$. Likewise, in the last practice block (block 10) there was significantly less mediated retrieval in the experimental condition than in the control condition, $t(29) = 3.7522$, $p < .001$. At these moments, in block 5 and 10, the same amount of time had passed since the encoding of the words in both conditions, but the words in the experimental condition had been retrieved more often than the words in the control condition.

The conditions were also compared after the same amount of retrieval, but after a different amount of time. The comparison between block 3 in the experimental condition and block 5 in the control condition, at which point the words in both conditions were retrieved for the third time, showed that there was no significant difference between the conditions after three retrievals, $t(29) = -0.17124$, $p = .87$. However, the comparison between block 4 of the experimental condition and block 10 of the control condition, at which point the words in both conditions were retrieved for the fourth time, showed that there was a significant difference between the conditions, $t(29) = -4.2461$, $p < .001$ after the fourth retrieval.

RQ2: Do shift differences between learners predict differences in performance later on?

Test performance. This section contains only the scores on the experimental words, which are the most informative since these have had considerably more retrievals than the control words. The performance of the participants is visualized in Figure 5. The score of the participants on the receptive test had an average proportion of correct responses of 90.5% ($SD = 12.1\%$). In this score, typing errors were controlled for (see method section). Figure 5 indicates the presence of a ceiling effect for performance on the receptive test. On the productive test two different margins were used to score the accuracy of the responses. Based on a Levenshtein Distance of ≤ 2 , in case the distance between the target and the response was small, the average proportion of correct responses on the productive performance was 79.5% ($SD = 13.8\%$). The average proportion of correct responses on the productive test based on a Levenshtein Distance of 0, in case the response had complete overlap with the target, was 51.2% ($SD = 20.3\%$).

Correlation between shift moment and performance. Table 5 summarizes the correlations between the moment of shift and the test performances on participant level. The Pearson correlation between the moment of shift and the receptive performance was not significant ($r(19) = -0.30$, $p = 0.19^4$). Likewise, the Pearson correlation between the shift moment and the productive performance, based on a Levenshtein Distance of ≤ 2 , was not significant ($r(19) = -0.34$, $p = 0.13$). However, the Pearson correlation between the moment of shift and the productive performance, based on a Levenshtein Distance of 0, was significant ($r(19) = -0.46$, $p = 0.03$). This indicates that participants who shifted earlier performed better on the productive test, based on a Levenshtein Distance of 0, than participants who shifted in a later block.

In addition, three influential cases were observed based on the Cook's Distance, which indicated the influential points that could have critical impact on the outcome of the analysis. Both Table 5 and Figure 5 display the correlations between the shift moment and the test performance including ($n_{pp} = 21$) and excluding ($n_{pp} = 18$) these three influential cases. Without these three influential cases, the Pearson correlation between the moment of shift and the performance on the receptive test was not significant ($r(16) = -0.37$, $p = 0.13$). Likewise, both the Pearson correlation on the productive tests were not significant, neither based on a Levenshtein Distance of ≤ 2 ($r(16) = -0.38$, $p = 0.12$) nor based on a Levenshtein Distance of 0 ($r(16) = -0.30$, $p = 0.22$).

⁴ df = number of observations (21 shifted participants) - 2

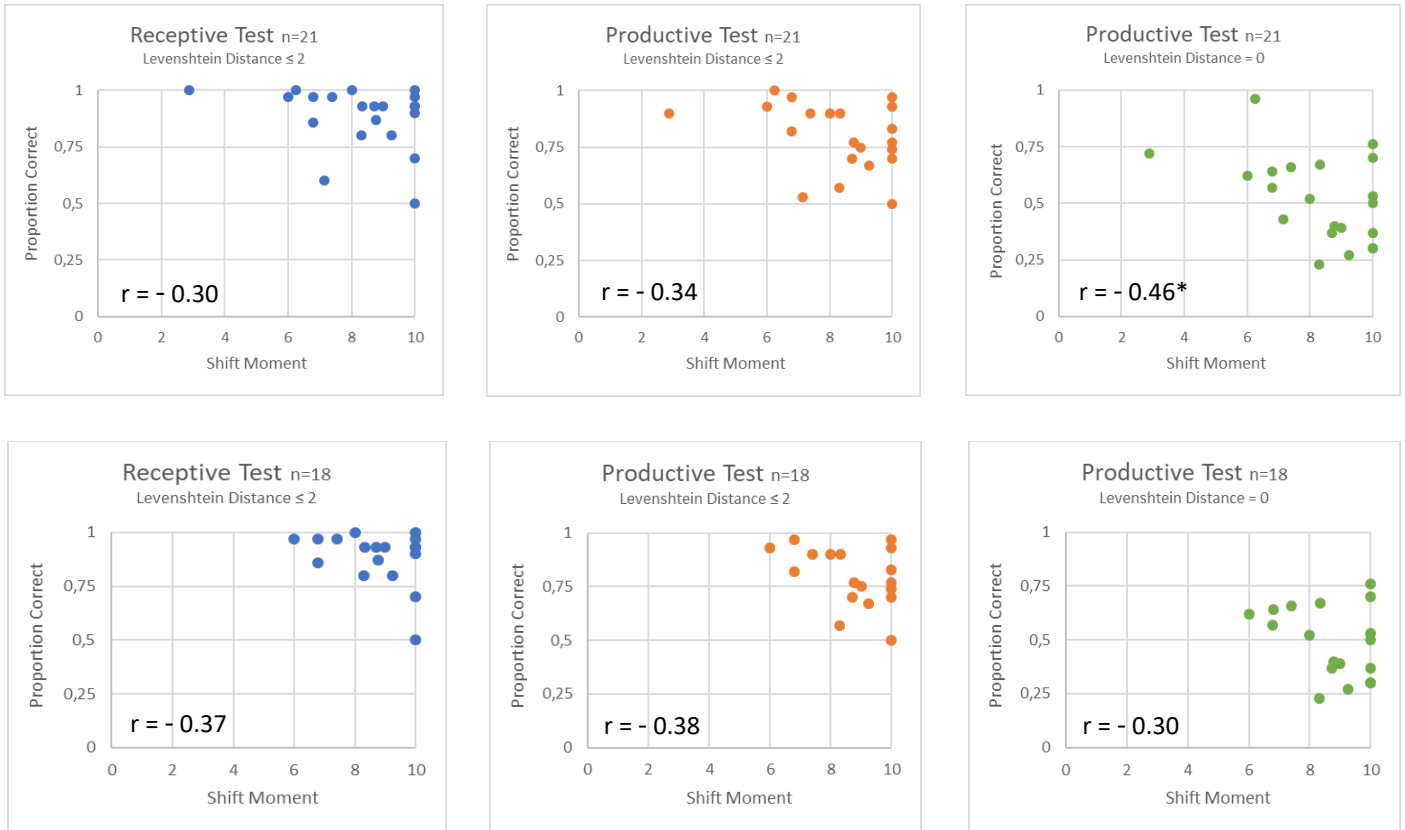


Figure 5. Correlation between the average shift moment and test performance at participant level, including ($n_{pp} = 21$) and excluding ($n_{pp} = 18$) potentially influential cases.

Table 5

Correlations between the average shift moment and performance on the tests at participant level, calculated for all shifted participants ($n_{pp} = 21$) and after exclusion of potentially influential cases ($n_{pp} = 18$).

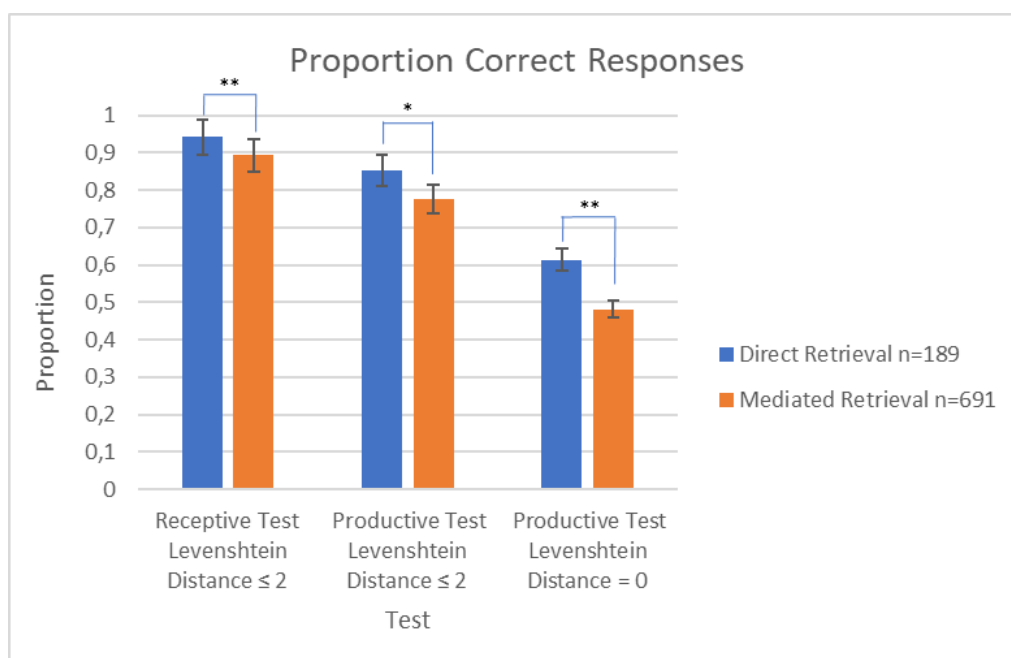
	All shifted participants				Without 3 potentially influential cases			
	t-value	df	Pearson correlation	p-value	t-value	df	Pearson correlation	p-value
Receptive test Levenshtein Distance ≤ 2	-1.359	19	-0.30	0.19	-1.6051	16	-0.37	0.13
Productive test Levenshtein Distance ≤ 2	-1.5647	19	-0.34	0.13	-1.6581	16	-0.38	0.12
Productive test Levenshtein Distance = 0	-2.2849	19	-0.46	0.03*	-1.2706	16	-0.30	0.22

RQ3: Do shift differences between words predict differences in performance later on?

Proportion correct. The test scores on words that were translated using direct retrieval from a certain point in the practice session (from now on abbreviated as *direct words*) were compared to the test scores on the words that were mediated throughout the whole practice session (from now on abbreviated as *mediated words*). Figure 6 demonstrates the proportion of correct responses on the different tests for both types of words (*direct* and *mediated*). This figure only includes words from the experimental condition.

The paired T-test comparison of the proportion of correct responses of each participant on the receptive test showed that the proportion of correct responses was significantly higher for direct words ($M = 94.2\%$, $SD = 23.5\%$) than for mediated words ($M = 89.3\%$, $SD = 30.9\%$), $t(20) = 2.8652$, $p = .01$. Likewise, on the leniently scored productive test, based on a Levenshtein Distance of ≤ 2 , the proportion of correct responses was significantly higher for the direct words ($M = 85.2\%$, $SD = 35.6\%$) than for the mediated words ($M = 77.7\%$, $SD = 41.6\%$), $t(20) = 2.5578$, $p = .02$. Similarly, when the accuracy was scored strictly on the productive test, based on a Levenshtein Distance of 0, the direct words ($M = 61.4\%$, $SD = 48.8\%$) had a significantly higher proportion correct responses than the mediated words ($M = 48.2\%$, $SD = 50.0\%$), $t(20) = 2.7262$, $p = .01$.

Shift to direct retrieval and later performance. Additionally, Logistic Mixed Effects Models were used to gain insight in the precise relation between the presence of a shift from mediated to direct retrieval and the performance on the tests and the moment and the test performances on word level. All analyses included both participants and words as random effects (unless stated otherwise). A summary of the results is displayed in Table 6. In addition, the predicted probabilities for a correct response based on different predictors are presented in Table 7, calculated from these models using the `plg` function in R (R Core Team, 2017).



Note: significance level * $p < .05$, ** $p = .01$

Figure 6. Proportion of correct responses on the receptive and productive tests, calculated across all experimental words for both direct and mediated retrieval. The error bars represent the SEM.

Experimental and control condition. First, both the words from the experimental and the control condition were included in the models, in order to explore the differences between the effects for both conditions on the different tests.

Receptive test. The model with the presence of a shift as predictor showed a significant main effect of the shift to direct retrieval on the receptive performance on a word ($\beta = 0.8767$, $Z(1465) = 2.911$, $p < .01$, $OR = 2.40$) with higher odds of a correct response for direct words than for mediated words. Further, the model with condition as predictor for the receptive performance showed a significant main effect of condition on the receptive performance of a word ($\beta = 0.4687$, $Z(1465) = 2.547$, $p = .01$, $OR = 1.60$) with higher odds for a correct response for words in the experimental condition than in the control condition. The model with both shift presence and condition as predictors for performance showed that there were no significant interaction effects between presence of a shift and the condition on the receptive test ($\beta = 0.8082$, $Z(1463) = 1.435$, $p = .15$, $OR = 2.40$). This suggests that the relation between the presence of a shift to direct retrieval and performance on the receptive test was similar for both the experimental and control condition.

An additional model was performed that included the moment of shift as a predictor of the performance. The model fit ANOVA between different possible models showed that there was no significant difference between a model that included both words and participants as random effects and a model that included only participants as random effect ($p = .20$), therefore the only included random effect was participants. However, no significant relation between the moment of shift and the performance on the receptive test was found ($\beta = 0.03516$, $Z(269) = 0.398$, $p = .69$, $OR = 1.04$).

Productive test. The model exploring the performance on the productive test scored leniently, with a Levenshtein Distance of ≤ 2 , showed that the main effect of the presence of a shift to direct retrieval was significant ($\beta = 0.7144$, $Z(1465) = 2.931$, $p < .01$, $OR = 2.04$) with higher odds for a correct response for words that shifted to direct retrieval than for words that remained mediated. However, the main effect of condition was not significant ($\beta = 0.1342$, $Z(1465) = 0.898$, $p = .37$, $OR = 1.04$). This indicates that there was no difference between the odds for a correct response for words in the experimental and the control condition on the leniently scored productive test. Also, no significant interaction effects were found between the presence of a shift and the condition on the receptive test ($\beta = 0.23734$, $Z(1463) = 0.519$, $p = .60$, $OR = 1.72$). Again, the additional analysis that included the moment of shift as a predictor of the performance did not show a significant effect of the shift moment on the performance ($\beta = 0.07144$, $Z(268) = 0.726$, $p = .47$, $OR = 1.07$).

The model exploring the performance on the productive test when scored strictly, with a Levenshtein Distance of 0, showed that the main effect of the presence of a shift to direct retrieval was significant ($\beta = 0.7778$, $Z(1465) = 3.927$, $p < .001$, $OR = 2.18$) with higher odds for a correct response for direct words than for mediated words. The main effect of condition was not significant ($\beta = 0.20838$, $Z(1465) = 1.633$, $p = .10$, $OR = 1.23$). This indicates that there was no difference between the odds for a correct response for words in the experimental and the control condition on the strictly scored productive test. Again, no significant interaction effects between the presence of a shift and the condition were found ($\beta = 0.55896$, $Z(1463) = 1.598$, $p = .11$, $OR = 1.48$). The additional analysis that included that shift moment as a predictor of the performance also did not show a significant effect of the shift moment on performance ($\beta = 0.04866$, $Z(268) = 0.796$, $p = .43$, $OR = 1.05$).

Experimental condition only. The previous analyses indicate that there are differences in the test performances between the conditions. In order to explore the exact effects of a shift from mediated to direct retrieval on the test performance, only the experimental words were included in the following models.

Receptive test. The model with the presence of a shift as predictor showed a significant main effect of the shift to direct retrieval on the receptive performance on a word in the experimental condition ($\beta = 1.1679$, $Z(876) = 2.621$, $p < .01$, $OR = 3.22$) with higher odds for a correct response for direct words than for mediated words. The model that included the moment of shift as predictor did not show a significant relation between the moment of shift and the performance on the receptive test in the experimental condition ($\beta = -0.2750$, $Z(186) = -1.628$, $p = .10$, $OR = 0.76$).

Productive test. For the leniently scored productive test, based on a Levenshtein Distance of ≤ 2 , the model with the presence of a shift as predictor showed a significant main effect of the presence of a shift to direct retrieval on the performance in the experimental condition ($\beta = 0.7823$, $Z(876) = 2.540$, $p = .01$, $OR = 2.19$) with higher odds for a correct response for direct words than for mediated words. The model that included the moment of shift as predictor did not show a significant relation between the moment of shift and the performance on the leniently scored productive test in the experimental condition ($\beta = -0.03369$, $Z(185) = -0.262$, $p = .79$, $OR = 0.97$). When scored strictly, based on a Levenshtein Distance of 0, the model that included the presence of a shift as predictor again showed a significant main effect of the presence of a shift to direct retrieval on the performance in the experimental condition ($\beta = 0.76528$, $Z(876) = 3.018$, $p < .01$, $OR = 2.15$). However, no significant effect was found in the model that included the moment of the shift to direct retrieval as predictor for the performance in the experimental condition ($\beta = -0.07420$, $Z(185) = -0.934$, $p = .35$, $OR = 0.93$).

Table 6

Summary of the results from the Logistic Mixed Effects Models

<i>Predictors</i>	Receptive Test	Productive Test	Productive Test
	Levensthein Distance ≤ 2	Levensthein Distance ≤ 2	Levensthein Distance = 0
Experimental and control condition			
Retrieval Type	<.01**	<.01**	<.001***
Condition	.01**	No Sig	No Sig
Retrieval Type & Condition	No Sig	No Sig	No Sig
Shift Moment	No Sig	No Sig	No Sig
Experimental condition only			
Retrieval Type	<.01**	.01**	<.01**
Shift Moment	No Sig	No Sig	No Sig

Table 7

Estimated probability of a correct response based on retrieval type and condition

	Receptive Test		Productive Test		Productive Test	
	Levensthein Distance ≤ 2		Levensthein Distance ≤ 2		Levensthein Distance = 0	
	%	β	%	β	%	β
Model with Retrieval Type as predictor						
Direct retrieval	97.3	0.8767	92.0	0.7144	65.9	0.7778
Mediated retrieval	93.7	2.7046	84.9	1.7270	47.1	-0.1179
Model with Condition as predictor						
Experimental condition	95.0	0.4687	86.9	0.1342	52.7	0.20838
Control condition	93.0	2.5900	85.3	1.7621	47.5	-0.09939
Model with Retrieval Type and Condition as predictor						
Direct retrieval & Experimental condition	98.1	0.8082	92.6	0.23734	66.6	0.55896
Direct retrieval & Control condition	95.6	0.3497	89.4	0.54374	56.6	0.39222
Mediated retrieval & Experimental condition	94.1	0.3062	85.2	0.05468	48.1	0.06205
Mediated retrieval & Control condition	93.7	2.5420	84.2	1.69678	45.7	-0.15327

Discussion

The present study investigated the thinking processes that underlie the keyword method for foreign vocabulary learning by means of the think aloud method. Specifically, it was investigated how the use of mediators changed over the course of retrieval practice and whether a possible shift from initially mediated to direct retrieval during practice influences the later retention of the words. The three main questions in this study concerned the exact moment of shift from mediated to direct retrieval, the relation between the moment of shift and the performance on participant level, and the relation between the shift moment and the performance on word level.

First, the exact moment of shift from mediated to direct retrieval was determined. It was expected that the mediation of the words would decrease after repeated practice and that a shift from initially mediated retrieval to direct retrieval would appear over the course of practice. More specifically, it was expected that this shift on average would appear between 5 and 10 repeated retrievals of the word based on an earlier study (Kole, & Healy, 2013). However, it was found that the retrieval of the vast majority of the words remained mediated throughout the entire experiment. Whereas previous studies suggest that a shift to direct retrieval is likely to

occur (e.g., Kole, & Healy, 2013; Rickard, & Bajic, 2003), the present study only found a shift on 21.6% of the words, even though the words were repeatedly retrieved for 10 times. There was, however, a decrease in mediated retrieval over the course of retrieval practice, with more and more direct retrieval in the later practice blocks. For those words for which a shift to direct retrieval occurred, the average moment of shift was after 8.27 retrievals, which is between 5 and 10 retrievals as expected.

The established decrease in mediated retrieval might support the direct model for memory retrieval (Adams, & McIntyre, 1967), which posits that the mediators merely serve as a learning context and that the mediated connection between the target form and its meaning will eventually change into a direct connection. The findings indicate that the retrieval of the words that were learned using the keyword method, changes into direct retrieval after practice. Thus, it seems that the retrieval of words that are learned using the keyword method might not forever remain mediated, but that it is likely that eventually a direct link is formed. Moreover, this possible change in memory retrieval after retrieval practice might support the identical elements model (Rickard, & Bajic, 2006), which states that retrieval practice leads to a fundamental change in the underlying memory representation, rather than in strengthening of the established links. In other words, instead of strengthening the mediated links, repeated retrieval practice seems to change these mediated links into direct links.

Additionally, it was found that the experimental and the control condition differed considerably in the proportion of words that were mediated. On the experimental words, which were retrieved more frequently than the control words, there was significantly more direct retrieval than on the control words. The two conditions differed only in the amount of retrieval they had on the words and not in the amount of time that had passed since the encoding of the words with keywords. Therefore, the difference between the conditions indicates that the decrease in mediated retrieval is an effect of retrieval and not purely of time and that repeated retrieval elicits the shift from initially mediated to direct retrieval.

The next step was to explore the relation between the moment of shift and the performance on both the receptive and the productive test on participant level, in order to answer the second research question. It was expected that an early shift from mediated to direct retrieval would lead to better test performances, because establishing a direct link is of fundamental importance to support and achieve efficient acquisition (Rickard, 1997; Rickard, & Bajic, 2006). Furthermore, a direct link might be less confusing than a mediated link which requires an extra step and activates additional and possibly irrelevant information that could lead to complications during retrieval (McElree, 2001; Rickard, & Bajic, 2003). The analysis of the test performance showed that most participants scored high, possibly performing at ceiling level on the receptive test and on the productive test the performance of the participants varied more. There was no correlation between the moment of shift and the correct responses on participant level, neither between the moment of shift and the test performance on the receptive test, nor between the moment of shift and the test performance on the leniently scored productive test. A significant correlation was only found between the moment of shift and the performance on the productive test when scored strictly, which suggests that the lack of correlations may be due to ceiling effects on the performance measures. However, this effect was not significant anymore after removal of three possibly influential cases.

Despite the lack of significance on these correlations, there does seem to be a trend in the correlations on participant level. The correlations namely all had negative values of moderate strength, which is in agreement with our predictions that an early shift leads to a better performance on the tests. Also, one significant negative correlation was found on the strictly scored productive test. This was the only test that was entirely free of ceiling effects. Possibly, the ceiling effects conceal certain effects of the shift moment on performance, but based on the

present findings these possible effects could not be revealed. Furthermore, the sample size may have been too small to show significance of these effects.

To answer the third research question, the relation between the actual presence of a shift from mediated to direct retrieval on the words and the test performance was explored. On word level, the effects of the shifts were investigated in two different ways. First, a distinction was made between words that shifted to direct retrieval and words that remained mediated. There was a relation between a shift from mediated to direct retrieval and the performance on the test. It was found that those words that had shifted, and therefore were directly retrieved during practice, were more likely to be translated correctly on both the receptive test and the productive test than the mediated words, on which no shift had occurred. Thus, the actual presence of a shift to direct retrieval during practice seems to be predictive for the later retention of the words, suggesting that a shift to direct retrieval might enhance the retention of the words. This effect of retrieval type was especially strong on the strictly scored productive test, which is the most informative test because it had no ceiling effects on performance. These findings support the hypothesis that direct retrieval is established in successful acquisition (Rickard, 1997) and that the direct, one step retrieval provides better retention of the words than mediated, multiple step retrieval (McElree, 2001; Rickard and Bajic, 2003).

However, as for the findings on participant level, mixed model analyses on word level showed no relation between the moment of shift and later test performance. Together these findings indicate that the occurrence of a shift during practice has an effect on the later performance, but the moment of shift might not affect the performance. Thus, the present study indicates that the establishment of a direct link is related to the later retention of the words, possibly regardless of the moment on which this shift occurs.

In addition, the analysis on word level demonstrates an effect of the condition on the retention of the words. The experimental words were more likely to be recalled correctly on the test than the words in the control condition. Furthermore, the conditions that differed in the amount of retrievals on the words, also differed in the amount of mediated retrieval that occurred during practice. The experimental words, which were retrieved more often, were more likely to be retrieved directly and accurately on the test than the control words. The control condition was included in the experiment to ensure that the results of the present study were effects of retrieval rather than time. Thus, based on the established differences between the conditions, the decrease in mediated retrieval throughout the experiment was presumably due to the repeated retrieval of the words that were learned with the keyword method, instead of the time that had passed since the encoding of the words.

Implications

The high performance of the participants in the present study support the effectiveness of a combination of the keyword method and retrieval practice. After about 90 minutes of practice, the participants performed at ceiling level on the receptive test after one week. On the test of productive knowledge, the performance was also rather good, even though the participants had not been practicing their productive performance at all during the practice phase of the experiment. These productive performances can be explained by the identical elements model (Rickard, & Bajic, 2006), which states that the memory representations of the practiced form-meaning associations are bidirectional. So even when these associations are only practiced in one direction (e.g., retrieving the translation of the foreign word), the combination is also strengthened the other way around (e.g., retrieving the foreign word that corresponds with the translation). However, this untrained direction is a bit weaker than the direction that has been repeatedly practiced.

These high performances support the proposed effectiveness of the combination of retrieval practice and the keyword method (Fritz et al., 2007) and are contrary to the doubts about the effectiveness of the keyword method on a longer term (as do Dunlosky et al., 2013). The findings in the present study indicate that the keyword method together with repeated retrieval practice might be an efficient strategy to learn the vocabulary of a foreign language. However, the effectiveness of the keyword method seems to increase when learners manage to shift from mediated to direct retrieval during practice.

The results on word level demonstrate the importance of the appearance of a shift from mediated to direct retrieval for later retention. Presumably a direct link between the target and its meaning improves the retention of the words. It may therefore be important that this direct link is established. The present study shows that this direct link can be successfully formed by means of the keyword method, provided that the newly learned words are retrieved from memory for a sufficient amount of repetitions. Based on our findings it is therefore recommended to retrieve the newly learned words for at least 8 times, because this seems to be the average moment on which a shift from mediated to direct retrieval appears.

Limitations

Several issues should be discussed that might have limited the present study. First, the absence of a shift on the majority of the words was inconsistent with the findings of earlier studies (e.g., Kole, & Healy, 2013; Rickard, & Bajic, 2003). One possible explanation is that the present study did not include enough retrievals in order to find the shift from mediated to direct retrieval and that there are more than 10 retrieval rounds needed. However, this explanation is in conflict with the findings of Kole and Healy (2013), who found that a shift from initial mediated to direct retrieval occurred between 5 and 10 retrievals of the words. Another possibility, is that the absence of a shift is partially due to our experimental settings. The think aloud method was used to investigate the thinking processes of the learners. Possibly, repeatedly saying aloud the keywords could have led to additional keyword practice. Repeating the keywords and mnemonics time after time could have resulted in more activation and rehearsal of the keywords, which could have led to more preservation of the keywords in the following trials. This might have influenced the development process of the word representations and might not reflect the thinking process as it would appear in a learning situation in which participants are not required to say their thoughts out loud.

Another possibility for these uniform responses could be that the participants got accustomed to saying aloud the keywords. This habit forming might have yielded a continued reference to the keywords in the think aloud data. In order to reduce these possible unnatural rehearsals and preservations of the keywords, further research could reduce the amount of trials in which the participants need to think aloud, by performing the think aloud task only in a limited number of blocks. So instead of 10 think aloud blocks, there could be 5 blocks of retrievals while thinking aloud and 5 blocks of retrievals in silence. However, it is unlikely that the only reason why learners kept using the mediators is the setup of the study. The control analysis showed that there was considerable more keyword use in the same practice blocks on the words that were retrieved less frequently. This indicates that the keyword use changes as a function of repeated retrieval practice. Thus, despite some possible limitations, there is support for the validity of the present findings, because they demonstrate that the retrieval genuinely changes over the course of repeated retrieval practice.

Another possible explanation for the constant reference to the keywords in the think aloud data is that the think aloud method sometimes might have yielded socially desirable answers. For instance, participants sometimes immediately said the translation and afterwards they quickly added the keyword and the complete mnemonic association sentence, as if they

had just remembered that they had to say this information out loud. Additionally, some participants indicated, during a small, oral debriefing after the experiment, that they said the keyword aloud in every trial because they thought that they were intended to do so. However, the instructions were neutral and the experimenter gave an additional instruction to “say everything that comes to mind, no more and no less”, in order to guarantee that the participants were not under the impression that they had to constantly repeat the keywords, even when they did not come to mind (see method section for a complete description of the instructions on thinking aloud). Despite these attempts to provide neutral instructions, some participants might have consistently said aloud the keywords because they thought they had to do so. It would be both interesting and advisable for further research to explore the effects of different instructions and then adjust the instructions for the think aloud measurements accordingly.

Furthermore, the results on participant level are calculated over a rather small dataset of 21 participants, since there were 9 participants that did not shift to mediated retrieval at all. This limits the statistical power of this study and the risk of spurious results based on the influential cases. Consequently, it is advisable to expand the current dataset and investigate the significance of the effects on participant level with a more powerful dataset. This possible limitation only applies for the analyses on participant level, since the analyses on word level were performed over an extensive dataset with a total of 1469 words.

Also, the comparison between the complete dataset and the correct dataset, that contained only the words that were recalled correctly throughout the entire practice phase, showed a difference between these two datasets in the experimental condition. For reasons of sample size and power, it was decided to focus on all retrieval attempts, both correct and incorrect, and to use the entire dataset for the analyses. Besides, these incorrect retrieval attempts elicited additional practice on the words, since the participants received feedback on their performance and were presented with the correct translation in case they gave an incorrect response. Therefore these incorrect retrieval attempts were included in the analyses. Additionally, it might be wise to explore the possible differences in the results on merely the correct subset of the experimental words as well. Even though the average shift moment of the complete dataset and the correct dataset did not seem to differ too much, the subset of correct retrievals might yield differences in the determination of the moments of shift. For instance, in case the retrieval of a word was direct in the block 4, 5, 6, 8, 9, and 10, but the word was incorrectly translated in block 7, the present study scored such incorrect retrievals as ‘not shifted’ and these were therefore treated as mediated retrievals. Hence, the shift pattern of this example would be *MMMDDDMDDD* and the shift moment would be determined at 7 retrievals. However, when only the correct retrievals were included, the shift pattern would be *MMMDDD-DDD* and the shift moment would be determined at 4 retrievals. It might be interesting for future research to explore the possible differences in shift behaviour between the two datasets, by taking the accuracy on each trial into account.

Another limitation is that the presence of a shift from mediated to direct retrieval was not manipulated in the present study, since the main focus of this study was on the moment of shift instead of the presence of a shift. Therefore, the conclusions about the presence of a shift are based on an observation of a relation between the shift presence and the performance on the tests. We can therefore not rule out the possibility that this relation is driven by other factors, like the difficulty of the words. In order to determine whether shifts genuinely affect the later retention, a different setup would be needed. For instance, it is possible to manipulate the difficulty of the words and explore whether easier words are shifted more often than difficult words.

Furthermore, the results on the ratings were not used as a determiner for exclusions for the analyses. This decision was made because the rating scores varied too much between the participants, therefore the ratings could not be used as a well-grounded indicator of possible

exclusions. However, it would be interesting to further examine these ratings, since they might give more insight in the explanations for the lack of correlations on participant level. In case a mnemonic is illogical or unusable for a participant, this could interfere in the accurate retention of these words. Although a correlation between the ratings and the test performance was not found in the present study, it could be advantageous to still exclude a certain proportion of words with a low rating. This should lead to different exclusions across the participants, since the ratings of the words varied considerably between participants.

Lastly, in the stimuli set we tried to control for as many confounding and confusing features as possible, since the keyword method is most effective when the stimuli are suitable and well thought through (Atkinson, 1975). One way in which we tried to achieve this, is that we only used keywords in Standard Dutch. For instance, in the dataset of the study of van den Broek et al. (in press), from which a considerable part of the Swahili words in the present study were selected, the Swahili word *duka* (store) was sometimes accompanied by a mnemonic with the Dutch word *doekoe*, which is Dutch slang for *money*. Since this is not a standard Dutch word, this keyword was replaced by *doek* (cloth), in attempt to ensure that the keyword was well-known by all participants. However, when inspecting the responses of the participants it became clear that some participants still used *doekoe* as keyword. This indicates that this word elicited an alternative keyword, which might interfere with the intended mnemonic association. Therefore, in addition to the words that were frequently confused, it might be convenient to replace *duka* with another Swahili word or to adjust the corresponding keyword.

Additional observations

The analyses in the present study were hypothesis driven. However, besides these main findings, several additional observations were made during the analysis of the think aloud data, which will be reported here. The most striking additional observation concerns the keyword use on the words that remained mediated throughout the entire experiment. Although the mediation did not disappear, the think aloud data indicated some kind of change over the course of retrieval practice on these mediated words. Initially, the mediated words were mostly translated while referring to the entire mnemonic association sentence. So, in the initial retrievals, the learners seem to use the complete mnemonic to get to the correct translation of the word. However, over the course of practice the reference to the entire mnemonic decreased and changed into a reference to the keyword only. This development was observed on 33.9% of the words that were mediated throughout the entire experiment. So in addition to the 18.6% of all words, both experimental and control, that shifted to direct retrieval, 33.9% shifted to a shorter form of mediated retrieval, and only 47.5% of the words were fully mediated throughout the entire practice phase. Possibly, this observation indicates that there might be an additional, intermediate step between completely mediated and direct retrieval.

Figure 7 shows three hypothetical retrieval routes to retrieve words that were learned using the keyword method, all differing in the number of steps they require to get to the correct translation. First, the figure illustrates a direct route in which the translation of the target is directly retrieved, in one step (the black route). Second, a limited mediation route is illustrated in which the retrieval of the target is mediated in a limited way (the blue route). This limited mediation route requires only one additional step to retrieve the translation of the target, namely only by selecting the keyword first. Third, a complete mediation route is illustrated in which the retrieval is completely mediated (the orange route). This complete mediation route requires two additional steps in order to get to the correct translation, namely by first selecting the keyword and then proceeding to the active visualization of the mental image. The focus of the present study, along with the focus of previous studies, was on the direct route (black) and the completely mediated route (orange), in which the complete mediation route and the limited

mediation route were taken together and treated as one. However, there might be a difference between these two mediation routes, which would be interesting to explore. The exploration of this possible distinction in retrieval routes, together with an exploration on the exact differences between the two sub routes of mediated retrieval in terms of retention might be an interesting topic for further research. It might give more insight in the precise underlying memory structure when learning the vocabulary of a foreign language by means of mnemonics.

A second interesting observation made during the inspection of the think aloud data was that participants did not always refer to the Swahili word. Instead, for some part of the words the participants referred directly to the keywords, without any reference to the Swahili word at all. This lack of references to the Swahili word appeared in 40.2% of all the trials in the experiment, irrespective of the type of retrieval (mediated or direct) of the words during practice. This gave the impression that on these words, the participants mainly focussed on the recognition of the keyword instead of on the exact form of the Swahili word. This would be interesting if true, because having only limited focus on the form of the novel word, might not be beneficial in vocabulary learning. A possible explanation for this observation would be that the word learning task that was used in the present experiment elicited this behaviour. Possibly, the participants focussed mainly on the recognition of the keyword because they knew that it was sufficient to successfully perform the retrieval task in the present experiment. Thus, this behaviour could be task-dependent, however it is also possible that in general learners focus more on the keywords than on the target words when learning vocabulary by means of the keyword method. This would be unfortunate if true, because it might lead to low performances when the learners have to produce the foreign words themselves. An interesting suggestion for a follow up study would be to investigate the references to the keywords and the target words, and whether there is a relation between the differences in focus on and the retention of the target words. This would be especially interesting to explore on a productive level, because in order to produce the Swahili words, the participants will need more information than only the information provided by the keywords.

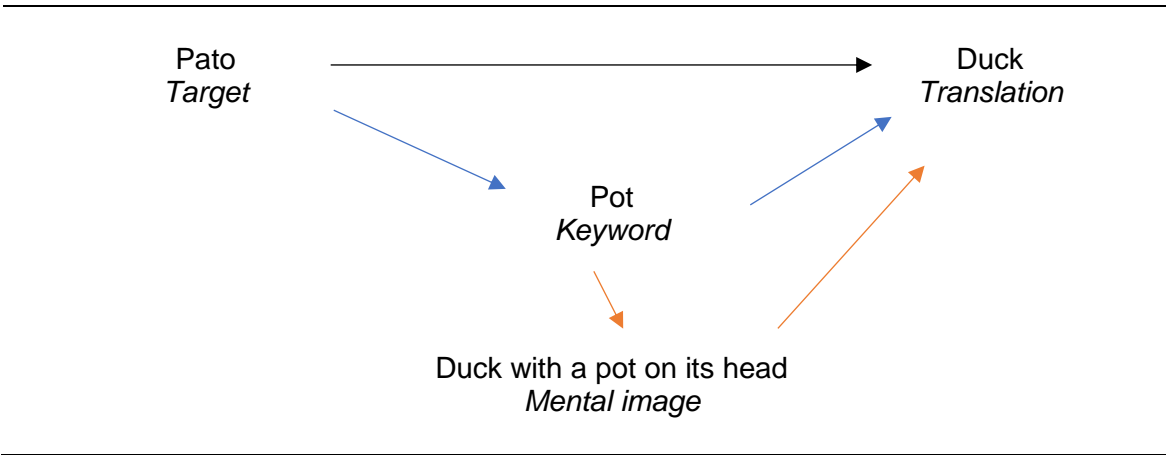


Figure 7. Possible retrieval routes: direct route (black), limited mediation route (blue), complete mediation route (orange).

Other interesting observations were made when exploring the different errors that were made in the practice phase of this experiment. There were some consistencies between the errors of different participants. Errors mostly occurred in the initial practice blocks and regularly on the same words. This indicates that the stimuli set contained a small number of stimuli that were more difficult for at least some part of the participants. However, this did not lead to persistent problems in the acquisition. The participants still performed at ceiling level, despite the possible confusions in the initial practice blocks.

For the majority of the errors the participant did not type in a response at all or there was no reference to the error in the think aloud data, and consequently the origin of these errors could not be evaluated (58%). The most frequent error was the confusion of two different keywords (21.9% of all errors). In this case the participant erroneously selected a keyword, that actually corresponded to a different Swahili word. As a result of selecting the wrong keyword, the participants also chose the wrong translation. For instance, the keywords corresponding with the Swahili words *pamba* and *panya* were often confused. The intended keyword for *pamba* was *panda* (panda) and the keyword for *panya* was *ranja* (syrup). However, due to the orthographic and phonologic overlap between these two Swahili words and keywords, the keywords were often mixed up. The same applies to the words *goti* and *godoro*, which were also frequently confused with one another, presumably because of orthographic and phonological similarities. Interestingly, it was observed that some participants even adjusted their mental image accordingly. The mental image of *pamba* (with *panda* (panda) as keyword and *wol* (wool) as translation) described a situation in which a panda lay in a pile of wool, and the mental image for *panya* (*ranja* (syrup) as keyword and *muis* (mouse) as translation) described a situation in which a bottle of syrup fell on a mouse. The mix up of the keywords led to the combination of these two mental images, resulting in a mental image of, for instance, a mouse in a pile of wool or a bottle of syrup falling on a panda, causing the participants to select the wrong translation. This type of error is specifically interesting because it highlights potential drawbacks of the keyword method, and demonstrates that the correct mental image did not come to mind and therefore has not been completely established yet. If the mental image was already accurately created and stored, it would not allow the confusion of two keywords since it interferes with the mental image.

Another different consistent error (7.7% of all errors) was the confusion of two stimuli, rather than only the keywords. For instance, the think aloud data showed that the Swahili words *mkate* and *paka* were often mixed up, since *mkate* had *kater* (tomcat) as keyword and the translation of *paka* is cat. So, the keyword of *mkate* is rather similar to the translation of *paka*, resulting in occasional confusion of these two stimuli across participants. Other errors were for instance that the participants accidentally typed in the keyword as a response instead of the translation, even when the think aloud data showed that they did think of the correct translation (5.5%), the response was an external word that was either semantically related to the keyword or the translation (3.7%), or the participants made incorrect deductions about the keyword or the translation based on the mnemonic association sentences (2.2%).

Conclusion

The purpose of the present study was to explore the use of keyword mediators over the course of retrieval practice. The results in the present study illustrate that the use of mediators decreases after repeated retrieval practice and that instead direct retrieval is increasingly used to translate the words. The average moment of shift from this initially mediated retrieval to direct retrieval appeared after about 8 repetitions. Further, the results of the present study show that the occurrence of a shift to direct retrieval during practice is a predictor of the later retention of the

words. Words that were retrieved directly during practice were more likely to be recalled correctly on both the receptive test and the productive test than words that remained mediated throughout the entire experiment. The moment of shift, however, was not predictive of later retention. Overall, it is concluded that the combination between the keyword method and retrieval practice is an effective strategy to learn the vocabulary of a foreign language. A shift from mediated to direct retrieval might make practice with keywords even more beneficial.

Literature

- Adams, J. A., & McIntyre, J. S. (1967). Natural language mediation and all-or-none learning. *Canadian Journal of Experimental Psychology*, 21(5), 436-449. doi: 10.1037/h0082995
- 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(5), 1063-1087. doi: 10.1037/0278-7393.20.5.1063
- Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of experimental psychology: learning, memory, and cognition*, 23(4), 932-945. doi: 10.1037/0278-7393.23.4.932
- Atkinson, R. C. (1975). Mnemotechnics in second-language learning. *American psychologist*, 30(8), 821-828. doi: 10.1037/h0077029
- Atkinson, R. C., & Wescourt, K. T. (1974). Some remarks on a theory of memory. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (pp. 485-498). London, England: Academic Press.
- Bower, G. H. (1972). Mental imagery and associative learning. In L.W. Gregg (Ed.), *Cognition in learning and memory* (pp. 51-88). New York: Wiley.
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36(2), 438-448. doi: 10.3758/MC.36.2.438
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6), 407-428. doi: 10.1037/0033-295X.82.6.407
- Crutcher, R. J., & Ericsson, K. A. (2000). The role of mediators in memory retrieval as a function of practice: Controlled mediation to direct access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1297-1317. doi: 10.1037/0278-7393.26.5.1297
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58. doi: 10.1177/1529100612453266
- Ellis, N., & Beaton, A. (1993a). Psycholinguistic determinants of foreign language vocabulary learning. *Language learning*, 43(4), 559-617. doi: 10.1111/j.1467-1770.1993.tb00627.x
- Ellis, N., & Beaton, A. (1993b). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *The Quarterly Journal of Experimental Psychology*, 46(3), 533-558. doi: 10.1080/14640749308401062
- Emmorey, K. D., & Fromkin, V. A. (1988). The mental lexicon. *Linguistics: the Cambridge survey*, 3, 124-149. doi: 10.1017/CBO9780511621062.006
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT press.
- Eriksson, J., Kalpouzos, G., & Nyberg, L. (2011). Rewiring the brain with repeated retrieval: A parametric fMRI study of the testing effect. *Neuroscience letters*, 505(1), 36-40. doi: 10.1016/j.neulet.2011.08.061
- Fritz, C. O., Morris, P. E., Acton, M., Voelkel, A. R., & Etkind, R. (2007). Comparing and combining retrieval practice and the keyword mnemonic for foreign vocabulary learning. *Applied Cognitive Psychology*, 21(4), 499-526. doi: 10.1002/acp.1287
- Hashimoto, T., Usui, N., Taira, M., & Kojima, S. (2011). Neural enhancement and attenuation induced by repetitive recall. *Neurobiology of learning and memory*, 96(2), 143-149. doi: 10.1016/j.nlm.2011.03.008

- Hulstijn, J. H. (1997). Mnemonic methods in foreign language vocabulary learning: theoretical considerations and pedagogical implications. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition: a rationale for pedagogy* (pp. 203-224). Cambridge: Cambridge University Press.
- Jääskeläinen, R. (2010). Think-aloud protocol. In Y. Gambier, & L. Van Doorslaer (Eds.), *Handbook of translation studies* (pp. 371-373). John Benjamins Publishing. doi: 10.1075/hts.1.thi1
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966-968. doi: 10.1126/science.1152408
- Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, 67(1), 17-29. doi: 10.1016/j.jml.2012.02.004
- Kole, J. A., & Healy, A. F. (2013). Is retrieval mediated after repeated testing?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(2), 462-472. doi: 10.1037/a0028880
- Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, 65(2), 85-97. doi: 10.1016/j.jml.2011.04.002
- Kumar, V. (2017). The think aloud method: Some concerns addressed. *Journal of Modern Languages*, 15(1), 13-25.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, 10(8). 707-710.
- Lowe, W. (1997, August). Meaning and the mental lexicon. In *Proceedings of the Fifteenth international joint conference on Artificial intelligence* (pp. 1092-1097). Morgan Kaufmann Publishers Inc.
- McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(3), 817-835. doi: 10.1037/0278-7393.27.3.817
- Nation, I. S. P. (2001). Teaching and explaining vocabulary. In I. S. P. Nation (Ed.), *Readings In Methodology* (pp. 144-151). Cambridge: University Press. doi: 10.1017/CBO9781139524759.005
- Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory*, 2(3), 325-335. doi: 10.1080/09658219408258951
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76(3), 241-263. doi: 10.1037/h0027272
- Piribabadi, A., & Rahmany, R. (2014). The Effect of the Keyword Method and Word-list Method Instruction on ESP Vocabulary Learning. *Journal of Language Teaching & Research*, 5(5). doi: 10.4304/jltr.5.5.1110-1115.
- Pirolli, P. L., & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of experimental psychology: Learning, memory, and cognition*, 11(1), 136-153. doi: 10.1037/0278-7393.11.1.136
- Pressley, M., Levin, J. R., & Delaney, H. D. (1982). The mnemonic keyword method. *Review of Educational Research*, 52(1), 61-91. doi: 10.3102/00346543052001061
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, 330(6002), 335-335. doi: 10.1126/science.1191465
- R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

- Richardson, J. T. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, 5(4), 597-614. doi: 10.3758/BF03208837
- Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, 126(3), 288-311. doi: 10.1037/0096-3445.126.3.288
- Rickard, T. C., & Bajic, D. (2003). Automatic mediation or absence of mediation? Commentary on Crutcher and Ericsson (2000). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1381-1386. doi: 10.1037/0278-7393.29.6.1381
- Rickard, T. C., & Bajic, D. (2006). Cued recall from image and sentence memory: A shift from episodic to identical elements representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 734-748. doi: 10.1037/0278-7393.32.4.734
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in cognitive sciences*, 15(1), 20-27. doi: 10.1016/j.tics.2010.09.003
- Safa, M. A., & Hamzavi, R. (2013). The Effect of Mnemonic Key Word Method on Vocabulary Learning and Long Term Retention. *Journal of English Language Teaching and Learning*, 5(12), 1-15.
- Sayyadi, M. A., & Javadi, M. A. (2014). The Study of Learning and Retention of Vocabulary Through Keyword Method among Iranian EFL Learners. *The Iranian EFL Journal*, 10(5), 117-128.
- Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language teaching research*, 12(3), 329-363. doi: 10.1177/1362168808089921
- Someren, M. V., Barnard, Y. F., & Sandberg, J. A. (1994). *The think aloud method: a practical approach to modelling cognitive processes*. London: Academic Press.
- van den Broek, G. S., Segers, E., Takashima, A., & Verhoeven, L. (2014). Do testing effects change over time? Insights from immediate and delayed retrieval speed. *Memory*, 22(7), 803-812. doi: 10.1080/09658211.2013.831455
- van den Broek, G. S., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *Neuroimage*, 78, 94-102. doi: 10.1016/j.neuroimage.2013.03.071
- van den Broek, G. S., Takashima, A., Segers, E., & Verhoeven, L. (in press). Contextual Richness and Word Learning: Context Enhances Comprehension but Retrieval Enhances Retention. *Language Learning*.
- Wang, A. Y., Thomas, M. H., & Ouellette, J. A. (1992). Keyword mnemonic and retention of second-language vocabulary words. *Journal of Educational Psychology*, 84(4), 520-528. doi: 10.1037/0022-0663.84.4.520
- Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, 11(6), 571-580. doi: 10.1080/09658210244000414
- Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., Hink, R. F., & Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of clinical and experimental neuropsychology*, 33(1), 101-111. doi: 10.1080/13803395.2010.493149

Appendix

CENTRE FOR LANGUAGE STUDIES

TOESTEMMINGSVERKLARING

Naam onderzoek:
Verantwoordelijke onderzoeker:

Verklaring deelnemer

Ik heb uitleg gekregen over het doel van het onderzoek. Ik heb vragen mogen stellen over het onderzoek. Ik neem vrijwillig aan het onderzoek deel. Ik begrijp dat ik op elk moment tijdens het onderzoek mag stoppen als ik dat wil. Ik begrijp hoe de gegevens van het onderzoek bewaard zullen worden en waarvoor ze gebruikt zullen worden. Ik stem in met deelname aan het onderzoek.

Naam: Geboortedatum:
Handtekening: Datum:

Verklaring uitvoerend onderzoeker

Ik verklaar dat ik de hierboven genoemde persoon juist heb geïnformeerd over het onderzoek en dat ik mij houd aan de richtlijnen voor onderzoekers zoals verwoord in het protocol van de Ethische Toetsingscommissie Geesteswetenschappen

Naam:
Handtekening: Datum:.....

TOESTEMMINGSVERKLARING

Naam onderzoek: Onderzoek Masterscriptie *The use of the keyword method in foreign vocabulary acquisition*

Verantwoordelijke onderzoeker: mw. dr. J. Klatter

Verklaring deelnemer

Ik heb uitleg gekregen over het doel van het onderzoek. Ik heb vragen mogen stellen over het onderzoek. Ik neem vrijwillig deel aan het onderzoek. Ik begrijp hoe de gegevens van het onderzoek bewaard zullen worden en waarvoor ze gebruikt zullen worden. Ik stem in met deelname aan het onderzoek. Daarbij geef ik toestemming om audio-opnamen van mij te maken voor dit onderzoek, deze opnamen op te slaan in een databank, volgens de geldende regels van de Radboud Universiteit Nijmegen, en beschikbaar te houden voor wetenschappelijk onderzoek.

Toestemming gebruik audio-opnamen

Ik geef toestemming om (s.v.p. aankruisen wat van toepassing is):

Ja Nee

deze audio-opnamen af te spelen op congressen en voor onderwijsdoeleinden.

deze audio-opnamen te verspreiden via Internet voor publiek gebruik.

overige opmerkingen:

Naam: Geboortedatum:.....

Handtekening:

Datum:.....

Verklaring uitvoerend onderzoeker (zie ommezijde)

Verklaring uitvoerend onderzoeker

Ik verklaar dat ik de hierboven genoemde persoon juist heb geïnformeerd over het onderzoek en dat ik mij houd aan de richtlijnen voor onderzoekers zoals verwoord in het protocol van de Ethische Toetsingscommissie Geesteswetenschappen.

Naam:

Handtekening: Datum:.....