Acquisition of foreign language vocabulary in the context of consolidation models

D. Krusko¹

¹Department of Informatics, Siberian Federal University, Krasnoyarsk, Russia Federation

Abstract—This paper reviews vocabulary learning evidence using concordance "standard model" of memory consolidation and a new approach, the "multiple memory trace" theory. Based on the findings of the experiment of the foreign language vocabulary learning with arbitrary time intervals between repetitions of training information and precise fixation of the measured parameters, a statistical analysis was conducted and identified statistical regularities of the vocabulary acquisition. This has been hypothesized by the influence of vocabulary item translation frequency for vocabulary memorizing. I reviewed the impact and opportunities of spacing repetitions used. In the experiment, two different techniques in the learning of vocabulary were used to adequately assess the strengths and weaknesses of each approach of memory consolidation. Overall, these findings indicate some discrepancies with standard model of memory consolidation, as well as the benefits of the approach is the combined use of both theories.

Keywords: memory consolidation, multiple trace theory, vocabulary learning

1. Introduction

There are a number of attempts to describe the process of forgetting and remembering information through computation models. Each of the models is based on a theory of a consolidation of information in memory and includes a set of assumptions about the optimal construction of the educational process. This article examines the assumption that multiple trace theory more accurately describes the process of acquiring vocabulary and in particular, that the recall of information from memory has a greater impact on memorization than passive repetition.

Multiple trace theory (MTT) was formulated and formed from standard consolidation theory based on evidence of patients with retrograde amnesia and damage of brain regions [9]. The main points of controversy between theories are viewed on the process to recall the recent and remote memories (including the involvement in this process of the hippocampal complex) and the fact that MTT postulates that different types of memory are encoding, storage and retrieval in different ways, while the standard theory does not differentiate.

In particular, the standard consolidation theory assumed that hippocampal complex (HC) engages in the process of storing episodic and semantic memory only to the stabilization of the information in the memory (consolidation period) and further the information may be retrieval from neocortical circuits without HC. Survey data with retrograde amnesia show that memory loss is significantly different for autobiographical, episodic and semantic memories. If the memory loss from patients with temporal lobe lesions for the details of autobiographical memory spreads from several years, several decades to total loss of autobiographical events then episodic and semantic memory are relatively preserved [6].

Most studies on the vocabulary learning describes experiments about acquiring of a foreign language vocabulary that study participants had not studied previously. Study of the effect of an early studied information on the acquisition of a new information given little attention and
the nature of this process is not clearly understood. MTT suggests that our memory about some episode or some item of the knowledge is a set of distributed and interconnected memory traces. Based on this, I put forward the hypothesis that in memorizing vocabulary items affects frequency of their translations in already studied language.

While most of the debate is on the role of hippocampus in the process of retrieving information from memory in order to support one or the other theory, almost no attention is paid to the functional features of both theories to practical use.

The purpose of this article is to statistically prove the standard consolidation theory and the multiple trace theory based on experimental data, the study of foreign language vocabulary. Quantify the impact of recall trials and passive study-only trials on memorization. The results can be used to change the approach to the study of foreign language vocabulary, and to refine the existing computational models of forgetting and remembering information.

2. The standard consolidation and multiple trace theories

Because multiple trace theory was formed from the standard consolidation theory due to the accumulation of contradictions, some points of view relating to the formation, maintenance and recovery of episodic memory are shared with the standard consolidation theory. The postulates of the theory put forward, which is reflected in this article. Common points in the two theories are [9].

- “The hippocampal complex rapidly (and obligatory) encodes all information that is attended or consciously apprehended. This process involves what is called short-term consolidation.
- This information is sparsely encoded in a distributed ensemble of hippocampal complex neurons”.

Interpretation of these items can serve a partition of the memorizing process for at least two stages. During the first stage all information are encoded (it is not necessary that the various types of information are processed in the same way) before the subsequent preservation. Naturally it is expected that after the information is sparsely encoded in a distributed ensemble of neurons, it begins to have quite different characteristics in speed of forgetting and the probability of retrieval from a memory. This assumption is reflected in the working memory models [1], [2].

The following postulates of MTT are different from the standard consolidation theory [9].

- “Each re-activation of memory trace occurs in an altered neuronal and experiential context.
- Because the hippocampal complex obligatorily encodes all information that is attended, the re-activation of a memory trace results in the creation of a newly encoded hippocampal trace, which also is sparse and distributed.
- By virtue of indexing a similar set of neocortical neurons encoding the features of the information in the memory, each such trace shares some or all of the information about the initial information”.

For example semantic information about the world as Grand Canyon located in Arizona, sushi is Japan food and etc, being acquired in some event during the consolidation process separated from the original event and stored independently.

In MTT, as a result of each successful retrieval of information for example in case exam will create new traces by replicating the recalled. If a memory has more than one trace, then the effect of all of them on memory retrieval is additive: any trace may cause retrieval, and the reliability of each trace action is independent of the number of traces. Therefore, failure in retrieval only occurs when all traces of a given memory fail. Recall of information in MTT is given a more important role in the process of remembering. At the same time the standard consolidation theory emphasizes the impact of rehearsal memorized information. With the amount of information rehearsals the strength of the information in memory increases and therefore it is less likely that this information will be forgotten.

On the basis of the postulates of MTT is reasonable to assume that the number of memory
traces for vocabulary items with a high occurrence frequency will be more, than for rare vocabulary items. Accordingly the probability to create a stable trace of foreign language vocabulary item in the memory should be higher for items whose translation has a high occurrence frequency, than for rare vocabulary items.

3. The Experiment

I performed an experiment on studying the vocabulary of a foreign language to test both of the above theories. In this experiment, 15 participants learned a set of 80 Russian-English word pairs. Studying of the vocabulary occurred within two months on the principle of always available e-learning system. That is participants were able to use the system at any time convenient to them and they worked with the system as long as they wanted. During the whole training process, the e-learning system collected anonymous statistical information about the learning process, as all the participants were informed before the experiment.

This experiment accomplished several goals. First, it provided a collection of extensive statistical information about the process of foreign vocabulary learning by means of online e-learning system in an informal and free atmosphere. To eliminate the effect of different algorithms the choice of words for lessons and exams, all participants worked with the Rastrigin’s adaptive model for education systems [10]. The choice of the model was based on prior results. In an earlier experiments this model showed one of the best results in comparison with the Bush-Mosteller, Krichevsky, Miller-McGill, Restla, Thurstone and Hull models [10]. A distinctive feature of Rastrigin’s model is that it assumes, that the process of memorizing different vocabulary items is not the same. Some elements of vocabulary can be learned with less difficulty and faster than others.

Second, the experiment provided a statistical test of both theories. The results showed which of the theories is the most complete and accurate description of the process of acquiring vocabulary. These results are particularly important when choosing a theory for the construction of an adequate computational model that can be used in the process of foreign language teaching. Experiment results allow the incorporation of additional components in existing computational models. For instance in the design of e-learning vocabulary system difference between recall-or-restudy trials and study-only trials (both trials techniques described below) may be taken into account in constructing the learning process.

Third, experimental data allow to test hypothesis about impact the frequency of vocabulary item translation on the memorizing vocabulary items process.

Fourth, the experiments provides evidence that allows visualization, in order to survey the possible regularities and investigate the possible patterns and nomination of assumptions, which may also be included in computational model. This evidence helps to understand long-term memory process better, which will aid in both model and theory development.

4. Experiment Design

During the experiment to study the vocabulary, two different approaches to learning have been tested. The e-learning system was both combined for two different approaches to learning: study-only approach and recall-or restudy approach. Study-only trials or passive study is when vocabulary elements in turn with their translation are displayed to the student and the student decides her/himself how long he/she will look to a particular vocabulary element, and when to move on the next. The recall-or-restudy trials were conducted by the type of an examination for the elements shown earlier in study-only trials. The essence of these approaches is described below.

E-learning system shows Russian word to the participant in the upper left side of screen and the participant had to recall its meaning. The participant was asked:

"Do You know this word? Yes/Maybe/No"

If the participant was confident in his knowledge of the vocabulary item, the system goes to the next vocabulary element of the set for the exam.
If the participant was not confident in his knowledge, he chose the "Maybe" element. In this case the vocabulary element was displayed with its translation and the participant was again asked: "Did you know it? Yes/No".

The answers of participants taking into account tips and time (accurate within 10 ms) were recorded. In order to validate the results, at end of the experiment, for each participant was conducted an examination on the entire list of words under the supervision of the experimenter. A participant had to tell translation of every vocabulary item to the experimenter. All experiment results of the participant that were untrue, that were untrue, were discarded and not included in the final selection for analysis.

The first presentation of each word was a study-only presentation of the pair. Each subsequent session consisted of two sets of words: one set for the recall-or-restudy trials and other for study-only trials. The number of elements in sets for each lesson are set individually at 1, .., 40 on uniform distribution law. Peculiarity of the Rastrigin’s model that words for study-only and for examination are selected based on likelihood of how well the experiment participant knows a specific word (for each participant is calculated according to its specific features). First select those words, which are less likely to be known by the user.

The stimuli were 80 Russian-English word pairs. Russian words were chosen so that words frequencies for their translations were distributed in the uniform distribution law. The data about word frequency were not used in the computa-
tional model in this experiment and did not influence at words choice as with study-only and recall-or-restudy trials. Word frequency were used only in the analysis process. The source of word frequency data was taken the corpus of contemporary American English (ANC). Russian words were taken from the A.P. Chekhov’s work of classical Russian literature. Word of the various parts of speech, of different lengths were selected for the experiment in their initial form (in the infinitive form for verbs). The main criteria for the choice of words was the requirements that the word should not originate from other languages, and that a set of words for the experiment must not contain paronymous words. As for the words translations that do not have an unambiguous translation were chosen to 3 English synonyms that best describes a word meaning in current context. Word pairs order of introduction and assignment to conditions was randomized individually for each participant. All study participants were students aged 25 to 33 and attended free of charge. There were 9 males and 6 females. All participants were fluent in English and have never learned Russian.

5. Results and Discussion

At the end of the experiment, after a preliminary analysis of the results, 9 participants data were selected. 5 were males and 4 were females. Some of the participants dropped out during the final exam for the reason that their previous answers did not correspond to the actual knowledge, other participants did not fulfill the exam. Our results are theoretically important because they strongly advocate which parameters of the learning process have an impact on the process of memorizing.

The method used conditional mutual information, which is reduction in the uncertainty between two random variables due to the knowledge of the third. For further processing the experimental data was aggregated by hours.

The conditional mutual information of random variables $X$ and $Y$ given $Z$ is defined by:

$$I(X; Y \mid Z) = \sum_{z \in Z} p(z) \sum_{y \in Y} \sum_{x \in X} p(x, y \mid z) \cdot \log \left( \frac{p(x, y \mid z)}{p(x \mid z) \cdot p(y \mid z)} \right)$$

For random variables $X$ and $Y$ were taken probability of success vocabulary element retrieval and count hours after last learning consistently. To find out which parameters influence mainly on the success of vocabulary element retrieval, a random variable $Z$ was calculated alternatively with different variants. Obtained in the calculation results are shown in Table 1.

<table>
<thead>
<tr>
<th>$Z$</th>
<th>$I(X; Y \mid Z)$, bit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count study-only trials</td>
<td>0.01</td>
</tr>
<tr>
<td>Count good answers on recall trials/Count recall trials</td>
<td>0.017</td>
</tr>
<tr>
<td>Count recall trials</td>
<td>0.03</td>
</tr>
<tr>
<td>Count good answers on recall trials</td>
<td>0.43</td>
</tr>
<tr>
<td>Count recall trials</td>
<td>0.43</td>
</tr>
<tr>
<td>Count study-only + recall trials</td>
<td>1.85</td>
</tr>
</tbody>
</table>

The table shows that recall trials have a much greater impact than the study-only trail. The calculation results demonstrate the greatest impact on vocabulary item retrieval is the sum of the count study-only trials plus count recall trials.

To check the dependency between participants’ answers, the number of hours since the last study, count study trials and count recall trials was performed in two-step procedure of the cluster analysis based on Bayesian Information Criterion (BIC). The results of the clustering procedure confirm evidence obtained with the conditional mutual information method. In Figure 2 it is seen that the number of recall trials has a greater impact than study trials. Another observation is that after some number of recall trials and study trials, an exam result in the short term becomes independent of the number of hours since the last the vocabulary item study.

To test the hypothesis that memorizing vocabulary items affects the frequency of their translations on the bases of the experimental data, the
relationship between vocabulary item translation frequency and the ratio of the number of correct answers to the number of wrong answers in the recall-or-restudy trials was plotted (see Figure 3). These evidences correspond to the hypothesis and show that participants make more correct answers for vocabulary items whose translation frequency is higher. A reduction of the correct answers ration at the end graphs (frequencies 16 and 38) may be associated with the mismatch interests of participants and a thematic of vocabulary items. To confirm the hypothesis it is necessary to conduct additional experiments with a large vocabulary and possibly thematic categorization of vocabulary items.

For the analysis of additional dependencies showed in Figure 1 relationships between the probability of success of vocabulary element retrieval, the number of hours since the last study (both study-only and recall-or-restudy) and the number of repetitions (the number of study-only element trials plus the number recall-or-restudy trials). For greater clarity, the data for visualization were limited by the number of repetitions to the value of 20. The probability of success of vocabulary element retrieval with the number of repetitions more than 20 tend to 1 independent of the number of hours since the last the vocabulary item study.

In Figure 1 it is seen in particular reducing the likelihood of the success of vocabulary element retrieval with some increase in the number of repetitions. These findings are consistent with the results of previous studies [3], [4]. In particular by comparing the experimental data in a verbal-learning paradigm with massed intervals and testing intervals was found that, spaced intervals are an advantage with respect to long-term memory [5]. Based on this findings the spacing effect was formulated. Further studies have found enhance in memory consolidation by distributed learning. Translating this effect on synaptic consolidation, it is assumed that mechanisms of the synaptic connections proliferation depends on intervals between memory reactivation. If the interval since the last memory reactivation is sufficient for the synthesis of proteins, long-term memory is
strengthen [7].

In Figure 1 it is also observed to extremes—minimum (in the region of 10-12 hours and 19-21 hours) passing further to extremes—maxima. I assume that these extremes—peaks are associated with periods of sleep participants. Memory consolidation during the sleep has received considerable attention in the standard theory. Experts claim that, the interval of single night will greatly increase the strength of the memory and presented the possibility that the the power of recollection undergoes a process of ripening and maturing during the time which intervenes [12].

These findings confirm the effect that recall trials are more effective than study-only trials [5]. Previous studies investigating spaced trials also demonstrate enhance long-term memory in comparison with massed trials [7]. At the same time, research into the effectiveness of the combined approach recall or restudy and study-only with optimal intervals for spaced repetitions is the subject of further studies.

These evidences generally suggest that the MTT more accurately describes the process of acquiring vocabulary. However, according to these data are, the sharing of the two theories gives better results than the use of any of the theories separately.

Acknowledgment

The research was conducted with the support of the joint program DAAD and Ministry of Education and Science of the Russian Federation "Mikhail Lomonosov".

References