Language Teaching Research

Spacing techniques in second language vocabulary acquisition: Short-term gains vs. long-term memory

Ulf Schuetze

The online version of this article can be found at:

Published by:

$SAGE

http://www.sagepublications.com

Additional services and information for *Language Teaching Research* can be found at:

Email Alerts: http://ltr.sagepub.com/cgi/alerts

Subscriptions: http://ltr.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://ltr.sagepub.com/content/early/2014/08/07/1362168814541726.refs.html

>> OnlineFirst Version of Record - Aug 8, 2014

What is This?

Spacing techniques in second language vocabulary acquisition: Short-term gains vs. long-term memory

Ulf Schuetze
University of Victoria, Canada

Abstract

This article reports the results of two experiments using the spacing technique (Leitner, 1972; Landauer & Bjork, 1978) in second language vocabulary acquisition. In the past, studies in this area have produced mixed results attempting to differentiate between massed, uniform and expanded intervals of spacing (Balota, Duchek, & Logan, 2007). A particular problem has been the point of testing that did not draw a clear line between short-term gains and long-term retention (Roediger & Karpicke, 2010). The experiments presented in this article addressed this issue. In the first experiment, 76 university students enrolled in a Beginning German class learned 24 content and 15 function words during a practice phase with a 'one plus three' design followed by three delayed post-tests. Results showed that in regards to short-term gains, the expanded group obtained higher mean scores than the uniform group, whereas in the long-term test it was the other way round. The second experiment used the same methodology with one exception: the practice phase was increased to a 'one plus four' design. Results confirmed those of the first experiment; in addition it was shown that function words are particularly difficult to recall for students using the expanded interval.

I Introduction

The study of vocabulary acquisition is often concerned with researching in what way associations, word-trees, and word-fields can be utilized to maximize the number of words students learn. Although most European languages have a repertoire of over

Corresponding author:
Ulf Schuetze, Department of Germanic and Slavic Studies, University of Victoria, PO Box 1700, Victoria, BC V8W 2Y2, Canada.
Email: ulfs@uvic.ca

100,000 words, knowing 2,500 words will give a student the ability to communicate effectively by either speaking or writing, as 2,500 words represent about 80 percent of the words used in everyday interaction (Schmitt, 2010). Therefore, in the context of teaching a second or foreign language at a college or university, the goal could be that students learn about 2,500 words. Judging by the way many textbooks are designed, this goal should be accomplished in the first two years of studying the target language. However, as some researchers report, it is surprising how little students remember at the end of a two-year university language program when it comes to vocabulary (Milton, 2009). Many of the words that had been learned in the textbook are forgotten quickly. For one, textbooks usually don't recycle words very well; that is, if they are introduced in one chapter they are not necessarily used again in subsequent chapters (Tschichold, 2012). In addition, little attention is paid to repeating words at the point of introduction (Nation, 2009; Zimmerman, 2010).

The last point is particularly interesting as the technique of 'spaced repetition' has been used to develop tools for teaching vocabulary as early as the 1970s (Leitner, 1972), even though the technique was not tested at the time but based on experience rather than experiments. However, around the same time, experiments testing memory were carried out in cognitive psychology. The first landmark study came from Landauer and Bjork (1978) who had students try to remember names of faces on prearranged cards. The same cards were repeated in a certain order with distraction cards in between, testing three intervals: massed spacing (no break in between each repetition), uniform spacing (regular break in between each repetition, for example two distraction cards), and expanded spacing (increasing break between each repetition, for example no distraction card, 3 distraction cards, 5 distraction cards). They found that the expanded schedule led to higher retention rates. That finding was accepted for a long time despite other studies in cognitive psychology (Balota, Duchek, Sergent-Marshall, & Roediger, 2006; Cull, 2000; Carpenter & DeLosch, 2005; Logan & Balota, 2008) that did not confirm those results. Reviewing studies using the spacing technique, Balota, Duchek and Logan (2007) as well as Roediger and Karpicke (2010) pointed out that the point of testing had varied from study to study, not drawing a clear line between short-term gains and long-term retention. This is a distinction that is being made by the model of working memory (Baddeley,1999, 2007; Baddeley & Hitch, 1974; Gathercole & Alloway, 2008), a model that has contributed much to our understanding of how phonemes and lexemes of words are processed in order to recall them at a later stage. In the context of that model, recalling words is seen as actively using them minutes (short-term gains) and days or weeks (long-term memory) after they have initially been processed. In second language acquisition, the need to recall words goes even further. The learner's goal is to recall words months after they have been learned. The question is how to accomplish that task; that is, to find out what spacing interval actually leads to long-term learning. Another question is if one interval works best for all words or if there is a difference as to how content and function words need to be treated.

II Context

Research on spacing has been trying to determine what type of intervals lead to the highest retention rates. This is an important question for second language learners as they can

study more efficiently if they know what interval to use in what time frame when studying vocabulary. Most of the studies carried out in cognitive psychology that I described in the previous section investigated 'spacing techniques' within one learning session, usually a single session of about 30 minutes. In second language acquisition, however, the pedagogy is often based on recycling material over longer periods of time. Therefore, analysing 'spaced repetition' over multiple learning sessions spread out over a week or two is of great interest.

As early as 1972, Leitner developed a prominent learning device called 'die Lernkartei' using the concept of structured cyclical repetition. Essentially, this was a memorization device consisting of flash cards and a box with 5–6 sections of progressively larger sizes. It was based on the principle of forgetting, using ideas from Pimsleur (1967), who in turn had used research by Ebbinghaus (1913), although neither of them carried out large-scale experiments. In self-experiments memorizing nonsense words in the first trial, Ebbinghaus measured the seconds he saved when relearning the words in each subsequent trial, which took place after 20 minutes, 1 hour, 9 hours, 1 day, 2 days, 6 days, and 31 days. He concluded that forgetting occurs within a few minutes after the first trial. The flash card boxes have been in use in the German school system for many years and their usefulness is promoted in teacher training programs. English as a second language textbooks publish sets of flash cards to match their corpora or create multimedia vocabulary learning environments based on cyclical learning. A general problem is that there is no control as to how much time a student spends reviewing a flashcard. Sometime in the 1990s, Mondria and Mondria-de Vries (1994) developed a computer-assisted learning program that was based on a textbook and incorporated the learning system of Leitner. Essentially, the program was a more modern version of the flash-card system featuring sound files and images. Yet again, no research was carried out. In particular, students had to decide for themselves if they had learned the words. Around the same time, Oxford (1990) promoted a staggered processing of learning material in her renowned textbook. Based on experience, she suggested seven encounters with optimal intervals of 15 minutes, 1 hour, 2 hours, 1 day, 4 days, 1 week, and 2 weeks.

In order to address the issue, Schuetze and Weimer-Stuckmann (2010) developed a study using an online platform called ViVo© (Virtual Vocabulary). The platform had several features. First, it used several modalities (illustration, audio, sample sentences, definition, L1 equivalents) to cater to the preferred learning style of each student (Kim & Gilman, 2008; Oxford, 2003; Rimrott, 2010). Second, prompted by the illustration, students had to type in a word learned every time they practiced, thereby fostering production in addition to comprehension. This feature was based on research by Hulstijn and Laufer (2001) as well as Webb (2005), who carried out many experiments on vocabulary learning arguing that words need to be actively practiced to get the attention they need. Third, students received immediate feedback (Nation, 2001). If they typed the word into the program incorrectly (correct spelling was checked by the program), they were asked to review all of the information for the word previously given, thereby receiving positive reinforcement. Fourth, words were systematically repeated via algorithms that had been programmed into the platform, testing a uniform as well as an expanded interval using a 'one plus four' design (initial encounter plus four practice sessions). Overall, 117 students of four sections of 'Beginning German I' at the University of Victoria, Canada participated in the study, divided into two groups. Results (Schuetze & Weimer-Stuckmann,

2011) showed that there were no statistically significant differences among students using the uniform or expanded interval on short-term gains (tests carried out four days as well as several weeks after the last practice). However, in a long-term test carried out several months after the last practice students using the uniform interval did obtain significantly higher scores than students using the expanded interval. The study showed that a clear distinction between short-term gains and long-term memory should be taken into account (Balota, Duchek, & Logan, 2007; Roediger & Karpicke, 2010) when testing spacing in vocabulary acquisition. Moreover, if second language learners are not to forget many words as soon as the semester is finished, as Milton (2009) reported they do, a uniform interval might be preferred. That, in turn, goes contrary to the common design of flashcard systems used today being based on Leitner (1972), which seems to be aiming at short-term gains.

In order to understand the differences between short-term gains and long-term memory, it is helpful to look at how words are processed when first encountered. Word processing takes place in working memory where two components are particularly important: the central executive that directs attention, and the phonological loop that codes, stores and rehearses incoming information (Baddeley,1999, 2007; Baddeley & Hitch, 1974; Gathercole & Alloway, 2008). The loop can code auditory as well as visual information. When a word is encountered and gets enough attention, its phonemes are processed and matched to a lexeme in Wernicke's area of speech comprehension. Additionally, its bound or system morphemes, if it has any, need to be identified. If the phonemes are matched to a lexeme, a record of the word is created. Another component of working memory is the so-called visual-spatial sketchpad that processes non-language information; for example, shapes we see, colors, location of objects, or movement. This information can also become part of the record of a word. When a word needs to be recalled, the record of the word is called up into Broca's area of speech production. The previously recorded information about the word assists in the process of matching phonemes to a lexeme by tagging the correct combination. Another component is the episodic buffer that provides a link between the executive, the loop and the sketchpad. The link is important as it allows information to be fast-tracked. Understanding this process, it becomes clear how the timing point of testing influences the retention. If words are tested minutes after the encounter, the loop is still active. At that point, it is difficult to say if a record has been created yet. In fact, that might depend on the word category, for example distinguishing between content and function words, as the processing of some words might be fast-tracked. What is then tested is if the loop is still active, but not if words have been committed to long-term memory. Therefore, studies using that type of testing have produced mixed results.

The study presented here served two purposes. First, it was designed to distinguish between short-term gains and long-term retention in the context of second language vocabulary acquisition using a uniform and expanded interval. Although Schuetze and Weimer-Stuckmann (2011) showed that the uniform interval might be preferred for long-term learning, these type of results need to be replicated in order to regard those results as reliable. Second, the study was designed to cater to the needs of students learning another language at the beginner's level. At that level, words are usually grouped by theme or grammatical category in textbooks or learning programs but no consideration is given to how words are actually processed in the brain.

A general distinction can be made between content and function words. This distinction is important in the everyday learning of a second language. Content and function words differ in the frequency of their occurrence (Aitchison, 2003; Milton, 2009), with content words being seemingly unlimited in number as they can be combined to make new words. Function words, on the contrary, are limited in number. Therefore, function words occur more often in any given written or spoken text because they bind content words into a meaningful sentence. In regards to processing, the main difference is that the lexeme of a content word (nouns, verbs, adjectives, some adverbs) expresses a rich concept and the word has bound or system morphemes. Function words, on the other hand, such as prepositions, conjunctions, fillers, articles and some adverbs have a simpler lexeme. Therefore, the lexeme is more difficult to distinguish from other lexemes because there are not as many unique features. This implies that function words need to be repeated more often in order to be processed to long-term memory than content words. Interestingly enough, as Schmitt (2010) points out, second language learners, in particular at the beginner's level, often report having difficulties learning function words.

Two other factors are important when it comes to processing words. If we come back to the model of working memory, several experiments going back to the 1960s and 1970s have shown that longer words (Baddeley, Thomson, & Buchanan, 1975) and phonologically similar words (Baddeley, 1966) are difficult to process. Every phoneme has a memory trace; that is, a time limit by which it has to be identified in the phonological loop in order to make room for other phonemes that need to be processed. When rehearsing the many phonemes of a longer word there might not be enough time for processing in the loop before all of them are identified, while rehearsing phonemes in similar sounding words makes it difficult to discriminate them sufficiently in the time available.

The study described in this article took those factors of processing into account. It was constructed to analyse how content and function words are processed when learning second language vocabulary and if that is related to the type of spacing. Furthermore, it distinguished between shorter and longer words as well as between phonologically similar and phonologically different words. It was hypothesized that due to the nature of word processing, participants who encountered these words for the first time would be able to recall more content words than function words and would have difficulties recalling longer and phonologically similar words. It was also hypothesized that the uniform interval would provide an advantage over the expanded interval when participants had to recall words in delayed post-tests on long-term retention.

Two experiments were carried out, both using the same methodology including the same number of words to be practiced and tested. However, the first experiment used a 'one plus three' design in terms of repeating words whereas the second experiment increased the repetition by one to a 'one plus four' design. The rationale was to investigate if the increase had an impact on the results. The 'one plus three' design was born out of necessity to accommodate the class schedule that term. The 'one plus four' design was in line with the previous study (Schuetze & Weimer-Stuckmann, 2011) to be able to compare results.

III Method

The first experiment was carried out in the fall of 2011 and the second experiment in the fall of 2012. In both experiments students who participated were enrolled in a course

called 'Beginning German I' at the University of Victoria, Canada. Each time, students were divided into two groups with one group using a uniform interval to acquire 24 content and 15 function words and one group using an expanded interval. Students listened and watched words on a screen. Their task was to copy each word on a piece of paper. This reflects how words are sometimes introduced by a teacher in class or how students are informed by a teacher how to study words at home.

1   Participants

Students were undergraduates age 17 to 24. Students taking 'Beginning German I' are supposed to have no or very little knowledge of German. However, in order to ensure that the students participating in the study were indeed beginners, students had to fill out a background questionnaire identifying previous knowledge of German, previous instruction in German, German heritage, proficiency in other languages, and providing demographic data on faculty, major, and years of study.

In the first experiment, of the 119 students enrolled in all four sections, 7 were excluded after the questionnaire had been analysed. Students used a 'one plus three' design, the schedule of which being tight. In addition, participants had their regular assignments to attend to. Those two factors contributed to the fact that of those 112 remaining students, only 76 participated in the experiment from start to finish. Students of sections one and two formed group number one using the uniform interval (33 undergraduates), and students of sections three and four formed group number two using the expanded interval (43 undergraduates).

In the second experiment students used a 'one plus four' design. Therefore only two of the four sections participated because the design did not fit into every section's schedule. Of 67 students, 5 were excluded (questionnaire) and of the remaining 62 students, 48 participated from start to finish, with section one using the uniform interval (24 undergraduates) and section four using the expanded interval (24 undergraduates).

2   Material

Thirty-nine words were selected for the study as this number of words is similar to what students have to learn actively in a time period of 8–10 days when enrolled in a German as a second language class at the University of Victoria. With active learning, I am referring to being able to use those words when speaking. At the beginning level, it is important to learn function and content words alike, even though there are many more content words in a language, in order to formulate complete sentences. In order to reflect the ratio of content and function words introduced in most German textbooks, 24 words were content words (subdivided into 8 nouns, 8 verbs and 8 adjectives) and 15 words were function words (subdivided into 5 prepositions, 5 conjunctions and 5 adverbs). In order to reflect the multiplicity of nouns, verbs and adjectives, two each were shorter words (two syllables or less), two were longer words (three syllables or more), two words that were not phonologically similar and two which were (for a complete list, see Appendix I). Word selection was based on frequency using Jones and Tschirner's dictionary (2006), although it should be mentioned that selection was not frequency driven but frequency informed. Word selection was also based on the textbook used for

'Beginning German I.' Only words not occurring in the first five chapters of that textbook were used to ensure that hearing and seeing the words with the PowerPoint slides was the only contact to the words participants had. This was particularly important for function words as they are limited in number.

3   Procedure

For each word a PowerPoint slide was created using the same font and same background color for all slides. On each slide the English word was presented on the left side of the screen. After two seconds the German equivalent word was presented on the right side and appeared for 6 seconds while the English word remained on the screen. The German word that appeared was also spoken. Each slide was shown for 8 seconds. The reason to include sound was to reflect the situation of how a learner often encounters a word in another language. The English word was presented first. There were two reasons to do so. One, in a learning environment where the target language is not predominantly spoken this is a commonly used approach. Two, in each of the three delayed post-tests, students were shown the English word and had to write down the German equivalent. Testing production is more challenging than simply identifying words, for example, showing the German word and asking for the English equivalent. However, to actively produce a word in a second language is what a learner is usually asked to do based on the assumption that learners want to communicate in that language.

Participants were asked to copy down the German word they saw and heard on a piece of paper. This was done as they had to write down the German words in the tests that followed so the activity of writing needed to be practiced. Each PowerPoint was presented in the last fifteen minutes of class time. Those students not participating in the study left the classroom at that point. At the end of class, all paper was collected by the research assistant to make sure that participants would not secretly study the words at home.

In experiment one, group one (sections one and two) followed a uniform spacing schedule: the PowerPoint was shown in a two–three–two interval; that is, it was shown on day 1, 4, 8 and 11. Ideally, a uniform schedule is 100 percent uniform, for example a two–two–two interval. However, the class schedule did not permit for that. Group two (sections three and four) followed an expanded spacing schedule: the PowerPoint was shown in a zero–one–three (PowerPoint on day 1, 2, 4 and 8) interval. Both groups saw and heard the words of the PowerPoint four times using the 'one plus three' design. Each time the order of the words was different to avoid testing effects; that is, the order of words primes the next word.

In experiment two in the following year, the uniform group (section one) viewed the PowerPoint on day 1, 4, 8, 11 and 15 whereas the expanded group (section four) viewed it on day 1, 2, 4, 8 and 15. Both groups saw and heard the words of the PowerPoint five times using the 'one plus four' design and each time the order of words was different.

4   Tests

Three post-tests were carried out after each experiment, one the day after the last practice (test 1), one four weeks after the last practice (test 2), and one 8 weeks after the last

practice (test 3). In all tests, the English word was given and participants had to write down the German word.

5   Limitations

In a classroom study it is difficult to control all factors that might influence how students perform. Although great care was taken in the selection of words for this study as well as in the procedure, such as collecting the lists of words students had written down after each session, there was the possibility that some students would be exposed to some of these words when studying at home. In my experience, having carried out similar studies in the past, the effect is balanced, that is, there are one or two students like that in each group. Since the study compared groups, this effect of extra exposure to the words acquired should not have had a major impact on the results.

Another limitation was that the uniform interval was not truly uniform due to scheduling issues as outlined above. However, the length of the study, in particular the second experiment where students had five encounters over 15 days, should have minimized this problem.

IV   Results

A multivariate analysis of variances was carried out on the 39 words practiced comparing the word categories (content words vs. function words) in relation to the interval (uniform vs. expanded). This was done for each test.

1   Experiment one

Test 1 was carried out one day after the last repetition. The average score of participants using the expanded interval was higher than of those participants using the uniform interval, although differences were not statistically significant ($F(1,74) = 1.784$, $p > .05$). In test 2, the average scores of students in both groups were similar, indicating that retention scores of the expanded group dropped more compared to the uniform group ($F(1,74) = 1.221$, $p > .05$). In test 3, both groups continued to drop their average scores, the expanded group a bit more than the uniform group, thereby widening the gap between them but again differences were not statistically significant ($F(1,74) = 1.000$, $p > .05$). In both groups the standard deviations were large, in particular in test 3, in relation to the average scores, indicating that some students did not remember any or only very few words (Table 1).

In all three tests of the uniform group, differences between content and function words were statistically significant (Test 1: $F(1,37) = 3.573$, $p < .05$) / Test 2: $F(1,37) = 3.058$, $p < .05$) / Test 3: $F(1,37) = 2.965$, $p < .05$). Students recalled content words significantly more than function words. The decreases in the group's scores from test 1 to test 2, as well as from test 2 to test 3, were similar across content and function words (Table 2).

In all three tests of the expanded group, differences between content and function words were also statistically significant (Test 1: $F(1,37) = 2.870$, $p < .05$) / Test 2: $F(1,37) = 2.676$ $p < .05$) / Test 3: $F(1,37) = 2.068$, $p < .05$). It was interesting to note that decreases

Table 1. Experiment 1: Tests 1, 2, 3 all words retention scores.

	U mean	SD	E mean	SD	F	Significance
Test 1	14.20	9.80	18.50	11.61	1.784	.189
Test 2	13.97	11.00	13.91	12.12	1.221	.641
Test 3	12.12	9.38	10.83	8.34	1.000	.386

Note. The Mean refers to the average score of all participants in each group (U = uniform interval group, 33 participants; E = expanded interval group, 43 participants).

Table 2. Experiment 1: Tests 1, 2, 3 uniform group content vs. function words.

	Content mean	SD	Function mean	SD	F	Significance
Test 1	17.66	10.16	8.44	5.98	3.573	.039
Test 2	16.80	10.42	7.44	5.10	3.058	.030
Test 3	15.66	10.12	6.22	4.56	2.965	.020

Note. In all tests, the number of participants was 33 in the group using the uniform interval. The Mean refers to the average score of all participants in the uniform group.

Table 3. Experiment 1: Tests 1, 2, 3 expanded group content vs. function words.

	Content mean	SD	Function mean	SD	F	Significance
Test 1	21.33	12.17	10.11	7.42	2.870	.040
Test 2	17.00	10.42	7.11	5.10	2.676	.019
Test 3	14.13	10.82	4.66	2.90	2.068	.015

Note. In all tests, the number of participants was 43 in the group using the expanded interval. The Mean refers to the average score of all participants in the expanded group.

in the group's scores from test to test were particularly high for function words. Logically, the decreases are more substantial because on short-term gains (test 1) the expanded group obtained higher mean scores than the uniform group. However, the long-term test (test 3) scores for content and function words were also both lower compared to the uniform group. The students using the uniform interval had lower scores on short-term gains to begin with, but then did not forget as many words in the long-run as those students using the expanded interval (Table 3).

Overall, it should to be noted that retention rates on long-term retention were in line with Milton's (2009) observation that many words are forgotten indeed. A possible explanation lies in the 'one plus three' design that was one repetition less than what had been shown in a previous study (Schuetze & Weimer-Stuckmann, 2011), which had revealed higher retention rates. The second experiment therefore increased the number of repetitions by one to a 'one plus four' design.

2   Experiment two

The pattern the analysis brought to light in experiment two was similar to that in experiment one. The expanded group had higher retention rates in test 1 ($F(1,46) = 1.840$, $p > .05$) but the uniform group in test 3 ($F(1,46) = 1.206$, $p > .05$). Interestingly, in test 2 the participants using the expanded interval obtained higher mean scores than participants using the uniform interval ($F(1,46) = 1.154$, $p > .05$). In experiment 1, test 2 scores had been similar. In both groups, standard deviations were not large indicating that there were not great fluctuations within the group as had been the case in experiment 1. In comparison to experiment 1, scores in all tests by both groups were higher. The extra repetition in the 'one plus four' design was beneficial to students in both groups. Taking into account the standard deviations in both experiments, it seems that having one more practice particularly helped students with lower scores to catch up. Furthermore, on short-term retention, students working with the expanded interval were able to recall words for several weeks before eventually the scores of these students dropped when tested 8 weeks after the last practice (Table 4).

In all three tests of the uniform group, differences between content and function words were statistically significant (Test 1: $F(1,37) = 5.605$, $p < .05$) / Test 2: $F(1,37) = 3.145$, $p < .05$) / Test 3: $F(1,37) = 3.954$, $p < .05$). Students recalled content words significantly more than function words. As in experiment 1, the decreases in the group's scores from test 1 to test 2, as well as from test 2 to test 3, were similar across content and function words (Table 5).

As in experiment 1, differences between content and function words were statistically significant across tests (Test 1: $F(1,37) = 5.915$, $p < .05$) / Test 2: $F(1,37) = 5.721$, $p < .05$)

Table 4. Experiment 2: Tests 1, 2, 3 all words retention scores.

	U mean	SD	E mean	SD	F	Significance
Test 1	17.43	7.05	21.90	7.49	1.840	.183
Test 2	15.97	6.88	18.01	6.70	1.206	.653
Test 3	14.42	6.43	12.01	6.12	1.154	.296

Note. The Mean refers to the average score of all participants in each group (U = uniform interval group, 24 participants; E = expanded interval group, 24 participants).

Table 5. Experiment 2: Tests 1, 2, 3 uniform group content vs. function words.

	Content mean	SD	Function mean	SD	F	Significance
Test 1	22.16	7.87	10.01	4.21	5.605	.022
Test 2	20.50	7.72	9.14	3.57	3.145	.013
Test 3	18.83	7.15	7.77	4.56	3.954	.008

Note. In all tests, the number of participants was 24 in the group using the uniform interval. The Mean refers to the average score of all participants in the uniform group.

Table 6. Experiment 2: Tests 1, 2, 3 expanded group content vs. function words.

	Content Mean	SD	Function Mean	SD	F	Significance
Test 1	23.13	8.34	13.76	6.63	5.915	.039
Test 2	21.96	8.40	9.11	3.92	5.721	.025
Test 3	18.00	7.79	6.26	2.84	3.870	.011

Note. In all tests, the number of participants was 24 in the group using the expanded interval. The Mean refers to the average score of all participants in the expanded group.

/ Test 3: $F(1,37) = 3.870$, $p < .05$). A closer look at function words indicated that students using the expanded interval seemed to forget those types of words quickly (Table 6).

In summary, repeating words four times (one plus four) led to higher recall rates than repeating them three times. In both experiments students using the expanded interval to repeat words for learning were more successful on short-term gains than students using the uniform interval, but on long-term retention, the uniform group obtained higher mean scores than the expanded group. This was consistent for content and function words. Overall, students of both groups recalled content words more successfully than function words. Retaining function words seem to be particularly difficult for learners using the expanded interval.

It also needs to be mentioned that within the content word category, longer and phonologically similar words were recalled the least, although the small number of words used did not allow for a statistical analysis. Although this indicates that processing these types of words is not only difficult in first-language acquisition, something Baddeley (1966) and Baddeley, Thomson, and Buchanan (1975) found a long time ago, but also in second language acquisition.

V Discussion

The results showed how important it is to distinguish between short-term gains and long-term memory, and to carefully choose the point of testing (Balota, Duchek, & Logan, 2007; Roediger & Karpicke, 2010). In both experiments, the expanded group did well on test 1 but eventually their scores started to drop. The uniform group was more consistent, starting with lower scores in test 1, but not forgetting as many words in the long-run as the expanded group and consequently having slightly higher scores in test 3.

It needs to be noted that none of the differences regarding the retention rates comparing both groups directly were statistically significant. Using a 'one plus four' design improved the rates, in particular for the expanded group, on short-term gains. In the literature on second language vocabulary there is some discussion about how often a word needs to be repeated. A survey by Schmitt (2008) showed that suggestions range from 3–20 repetitions. Naturally this depends on the word itself, as this study has shown, as well as on the learning situation, for example the level of proficiency and the learning environment. In the context of undergraduate students speaking English as an L1 acquiring a Germanic language at college or university at the beginner's level it seems that five encounters would be a good way to start.

In regards to processing, the results supported the hypothesis that learners recall more content words than function words and have difficulties recalling longer and phonological similar words. This was consistent using both, the uniform and the expanded interval. The results support our understanding of the phonological loop (Baddeley, 1999, 2007; Baddeley & Hitch, 1974; Gathercole & Alloway, 2008), as content words carry more information in terms of their lexemes and can therefore be discriminated using less repetitions than function words. The results give support to Schmitt's (2010) observation that second language learners report having difficulties learning function words. For teaching and learning, it might be helpful to not just follow suggestions made in textbooks or language programs that group words by theme or grammatical category, but to pay attention to the underlying factors of processing a word, such as the division between content and function words. If words are presented in textbooks or language programs by theme or grammatical category, the ones that are more difficult to process could be marked which might give students a better chance to acquire words in the long-run, repeating those marked words more often and using a uniform interval.

Furthermore, the results of the previous spacing study by Schuetze and Weimer-Stuckmann (2011) using the online platform ViVo© (Virtual Vocabulary) that provided a context for each word using a 'one plus four' design, were replicated with some variation. On short-term retention, there were no statistically significant differences between the uniform and the expanded group, but the expanded group did outperform the uniform group. On long-term retention, tests being carried out 8 weeks after the last practice, did not show statistically significant differences, either, although the trend favored the uniform group. The statistically significant differences that had been found in the 2011 study had occurred several months after the last practice.

The question then is what these results imply for the learner. When choosing one interval over the other, one might want to think about the learning goal when acquiring all these new words. If the goal is to learn words quickly in order to go on a vacation, for example, the expanded interval might lead to some success. If the goal is to study a language for a longer period of time, taking another course the following semester, it might be worth thinking about embracing the uniform interval. Most interesting in this context were the differences between content and function words. It seems that function words in particular are subject to a high rate of forgetting using the expanded interval, and are not retained all that much to begin with using either interval. At the same time, knowing those function words to formulate complete sentences is an important feature when learning another language. It might be helpful to use the uniform interval when acquiring these types of words and extending the number of repetitions beyond five. In addition, those words might be studied in the context of a phrase. As function words bind content words to a sentence, sometimes function words are presented in textbooks or learning programs by providing example sentences. The preposition 'to', for example, has several equivalents in German such as *zu, nach, bis, an* or *auf*. If an example sentence for each of those equivalents is provided, the process of acquiring that preposition is different than that of a content word, as the meaning of the preposition changes with the context. Although content words are also subject to having multiple meanings, they are perceived as independent due to the rich lexeme they carry (Aitchison, 2003). As a practical application it

might then be helpful to develop flashcard systems that are somewhat more sophisticated than the standard design that is based on expanding intervals providing discrete lexical items. These modified types of flashcard systems would apply both intervals, depending on the learner's goal as well as the type of word to be acquired. As an option they would provide sample sentences where necessary. Naturally, in view of the task of learning many new words in a relatively short period of time in a beginning second language class, the challenge is always to find a balance between number of words acquired and time spent to do so.

There are more questions to be answered in future research. One of them is to find out more about the rate of forgetting, as it was more evident in the expanded than the uniform group. A study could be designed to have students learn words until all of them have been acquired and then measure the rate of attrition in a series of post-tests. However, this type of study would likely have to be carried out outside the classroom as it poses some challenges to fit into a teaching schedule. A variation of that kind of study could be to use the methodology presented here but with less words. Having to practice less words might increase the recall rates on short-term gains, and if subsequently those rates are high, there is a possibility that long-term rates are higher as well as learners have a larger pool of words to draw from and consequently to forget.

Declaration of conflicting interests

Funding

References

Aitchison, J. (2003). *Words in the mind: An introduction to the mental lexicon*. Oxford: Blackwell.

Baddeley, A.D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, *18*, 302–309.

Baddeley, A.D. (1999). *Essentials of human memory*. Hove: Psychology Press.

Baddeley, A.D. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory: Volume 8* (pp. 47–89). New York: Academic Press.

Baddeley, A.D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour*, *14*, 575–589.

Balota, D., Duchek, J., & Logan, J. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. In J. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 83–106). New York: Psychology Press.

Balota, D.A., Duchek, J.M., Sergent-Marshall, S.D., & Roediger, H.L. (2006). Does expanded retrieval produce benefits over equal-interval spacing? Explorations of spacing effects in healthy aging and early stage Alzheimer's disease. *Psychology and Aging*, *21*, 19–31.

Carpenter, S.K., & DeLosch, E.L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, *19*, 619–636.

Cull, W. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*, 215–235.

Ebbinghaus, H. (1913). *Grundzüge der Psychologie* [Fundamentals of psychology]. Leipzig: Viet.

Gathercole, S.E., & Alloway, T.P. (2008). *Working memory and learning: A practical guide for teachers*. London: Sage Publications.

Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*, 539–558.

Jones, R.L., & Tschirner, E. (2006). *A frequency dictionary of German*. London: Routledge.

Kim, D., & Gilman, D. (2008). Effects of text, audio, and graphic aids in multimedia instruction for vocabulary learning. *Educational Technology and Society*, *11*, 114–126.

Landauer, T., & Bjork, R. (1978). Optimal rehearsal patterns and name learning. In M. Grüneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.

Leitner, S. (1972). *So lernt man Lernen: Der Weg zum Erfolg* [Learning to learn: The road to success]. Freiburg: Herder.

Logan, J.M., & Balota, D.A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, and Cognition*, *15*, 257–280.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Mondria, J., & Mondria-De Vries, S. (1994). Efficiently memorizing words with the help of word cards and 'hand computer': Theory and applications. *System*, *22*, 47–57.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I.S.P. (2009). *Teaching vocabulary: Strategies and techniques*. Boston, MA: Heinle.

Oxford, R.L. (1990). *Language learning strategies: What every teacher should know*. Boston, MA: Heinle.

Oxford, R. (2003). Language learning styles and strategies: Concepts and relationships. *International Review of Applied Linguistics in Language Teaching*, *41*, 271–278.

Pimsleur, P. (1967). A memory schedule. *The Modern Language Journal*, *51*, 73–75.

Rimrott, A. (2010). Computer-assisted vocabulary learning: Multimedia annotations, word concreteness and individualized instruction. Unpublished dissertation, Simon Fraser University, Burnaby, BC, Canada.

Roediger, H.L., & Karpicke, J.D. (2010). Intricacies of spaced retrieval: A resolution. In A.S. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork* (pp. 1–36). New York: Psychology Press.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.

Schuetze, U., & Weimer-Stuckmann, G. (2010). Virtual vocabulary: Research and learning in lexical processing. *CALICO Journal*, *27*, 517–528.

Schuetze, U., & Weimer-Stuckmann, G. (2011). Retention in SLA Processing. *CALICO Journal*, *28*, 460–472.

Tschichold, C. (2012). French vocabulary in Encore Tricolore: Do pupils have a chance? *The Language Learning Journal*, *40*, 7–19.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*, 33–52.

Zimmermann, C.B. (2009). *Word knowledge: A vocabulary teacher's handbook*. Oxford: Oxford University Press.

Appendix I

Word list

Part 1: Content words (nouns / verbs / adjectives): 24 total

Nouns:
der Turm, das Dach [the tower, the roof]
der Staatsangehörige, das Sonnenschutzmittel [the national, the sunscreen lotion]
die Liebe, die Niederlage [the love, the defeat]
der Berg, die Burg [the mountain, the fortress]

Verbs:
malen, trocknen [to paint, to dry]
beabsichtigen, vervollständigen [to intend, to complete]
freuen, weinen [to be happy, to cry]
rauchen, saufen [to smoke, to drink]

Adjectives:
schnell, tief [fast, deep]
rätselhaft, notwendig [mysterious, necessary]
elend, verrückt [miserable, crazy]
kühl, schwül [cool, humid]

Part 2: Function words (adverbs / conjunctions / prepositions): 15 total

Adverbs:
vielleicht, eigentlich, schon, wirklich, meist [perhaps, actually, already, really, mostly]

Conjunctions:
sondern, nachdem, deshalb, denn, obwohl [but, after, thus, because, although]

Prepositions:
seit, zwischen, mit, auf, jenseits [since, between, with, on, beyond]

Author biography

Ulf Schuetze is an Associate Professor of Second Language Acquisition. He concentrates on two areas: investigating questions of sustainability using information technology in second language learning and teaching, and analysing lexical processing from first encounter to long-term memory when acquiring and speaking another language.