How to create a vocabulary learning file with Excel

Markus Rude
Tsukuba University

Reference data:

Rude, M. (2006). How to create a vocabulary learning file with Excel.
In K. Bradford-Watts, C. Ikeguchi, & M. Swanson (Eds.) *JALT2005 Conference Proceedings*. Tokyo: JALT.

The acquisition of the lexicon is a central issue in learning a foreign language. In this paper, a technique on how to study vocabulary systematically by using a spreadsheet program such as Microsoft Excel will be introduced. In the theoretical part, the principle underlying Leitner card file system and how it can be set up on a computer will be explained. In the tutorial part, the step-by-step creation of a learning file will be discussed. As demonstrated in the presentation where participants with laptops were able to create their own file, this paper contains essential information needed by a reader to create the file system.

外国語学習において、語彙の習得は大切である。この論文では、エクセルのシートを使ってのシステマティックな語彙の学習法を紹介する。Leitnerカードファイルシステムについて基礎的な説明と、コンピュータでどのように使用するか述べる。また学習ファイルの作成についても説明する。ワークショップではノートパソコンを持って来た人は自分のファイルを作った。従って、読者はこの論文の情報で自分のファイルを作ることができるだろう。

The acquisition of vocabulary is a central issue in learning a foreign language. However, many learners lack sufficient active as well as passive knowledge of words and expressions. How to study vocabulary most efficiently is a debatable question. Words are believed to form a complex interconnected mental network (Meara, 2004). Nevertheless, many leading vocabulary researchers (Laufer, Meara, & Nation, 2005) believe, that the explicit study of vocabulary items by word cards or by computer files is still a useful element in the acquisition process of vocabulary.

In 1972, the German, Sebastian Leitner, introduced his Leitner card file system that helps a language learner to repeat vocabulary items in a systematic way. In the first part of this paper, the *Theoretical Part*, the design of Leitner's card file system and its learning method will be introduced. The reordering process of flashcards is done in such a way that vocabulary items are repeated neither too late nor too early: Cards, which a learner has forgotten, will be repeated more frequently, compared to cards, which the learner

could recall. It will be explained how this system can be realized on a computer using a spreadsheet program such as Microsoft Excel. Finally, a learning cycle using this file system is illustrated through a sequence of tables.

In the second part of this paper, the Tutorial Part, how to create a learning file with Excel will be explained systematically. In particular, readers with little or no experience with Excel are encouraged to try it on their PCs by themselves.

This paper is an adapted and extended version of an earlier report (Rude, 2004). In particular, the *Tutorial Part* is completely new and reflects the practical part of the workshop: The feedback provided by the workshop participants were useful in the modifications of the learning file and helped improve this paper considerably.

Why Excel?

What are the merits of using a spreadsheet program such as Excel for studying vocabulary? Of course, there are other more sophisticated software solutions based on Leitner's idea - like the one by Raedle (2005). However, there are several benefits of the proposed Excel solution:

- Almost everybody has access to a spreadsheet program like Excel (such as Microsoft Excel or a Linux version),
- Almost all students and teachers know to some extent how to use Excel, and
- The basic learning file can easily be adapted and extended.

In fact, some participants of the workshop have already tailored the Excel file according to their needs and thus provided input to the subsection *"Effective Hints ..."* in the *Tutorial Part*.

Theoretical background

Hermann Ebbinghaus

Ebbinghaus (1850-1909) was a memory researcher, two centuries ago, who came up with some interesting findings about forgetting. Best known of his results is what is today called "the forgetting curve" (Straub, 1997): If we learn something new, within just 20 minutes we will already have forgotten about 40% of it. The percentage of what is retained decreases with time over the next hour, the remainder being about 50%. This decrease continues the next day, with the remainder being about 30%. The next month it falls to about 20%. This curve slightly resembles the slopes of Mt. Fuji. Scientifically, this indicates that memory decreases more or less exponentially). The question must be this: If we forget everything eventually, why do we study at all?

Luckily, the forgetting curve never reaches 0%. If 20% of Ebbinghaus' nonsense syllables (these were his study material) were retained for as long as one month, we can reasonably expect at least the same retention for some meaningful learning material, such as vocabulary.

Sebastian Leitner

The German, Sebastian Leitner, concluded that about one out of five words learned will directly go to our long-term memory (corresponding to Ebbinghaus' 20% remaining

memory after one month). His proposition is "Repeat only the four that will be forgotten. Do not repeat the fifth, since it is already in the long-term memory; do not waste your time". However, the question is, how can we know, which is the ONE among the five?

In the 1970s, Leitner developed his "learning machine", which consists of a box with five sections of 1cm, 2cm, 5cm, 8cm and 14cm width (Leitner, 1972). They have respectively room for about 30, 60, 150, 250 and 500 cards. A new item on a flashcard is always put into section 1. If we feel like studying or if section 1 overflows, we take some or all of its cards and perform a self-test as follows:

- If we read a flashcard's front side and we cannot recall its rear side, we do not worry and just insert it at the back of section 1.
- If we can recall it, the self-test was successful and we put the card away into the back of section 2.

We can go on writing new cards and putting them into section 1. If it gets full again, we do the same self-test as before and thus thin it out.

As section 2 gets full, we take a bundle of cards from the front - about 10 - and perform the same self-test:

- If we do not know a card, we return it to the back of section 1.
- If we know it, we may put this "successful" card away into the back of section 3.

As section 3 gets full, we start the self-test with the front 10 cards: A card will either return to section 1, or proceed to the back of section 4, and so forth.

The increasing sizes in the sections determine increasing time intervals, after which each flashcard is repeated. Depending on frequency of study, these time-intervals range from just a few minutes or hours in section 1 up to several weeks or months in section 5. Leitner assumed that a card, which was successfully repeated 5 times in a row, had reached long-term memory and the information will never be forgotten.

The Excel file as a digital learning machine

Today, flashcards are out, and computers are in. We can therefore adapt Leitner's idea to a spreadsheet program, such as Excel. Assume a learner wants to study the Japanese equivalent of an English word. On the flashcard's front side he or she would write the English word (the trigger), on the back the Japanese translation of that word (the piece of information to be recalled from memory).

We proceed with an Excel worksheet as follows (Table 1): The cells of column A contain the English "trigger" (the content of the flashcards' front sides). Column B includes the Japanese words to be recalled (indicated in the rear sides of the flashcards). Column C shows the section numbers. A new word is always put into section 1. Each line of the Excel file now represents one flashcard of the learning machine.

As an illustrative example, we assume that sections 1 and 2 are completely filled with cards. Since a computer file never gets full, we have to limit the capacity of a section by a number: let us say we limit section 1, 2 and 3 to 3, 6 and 15 lines, respectively. This downscaling to 10% of the values above is only for the sake of illustration. Let us now simulate

the input of a new word and the associated self-test:

Table 1 is the starting point of our example: We want to learn the new word "arrangement" and insert it into section 1. However, section 1 is already full (section 1 already contains 3 items) and indeed overflows with the 4th entry (Table 2). Through two self-tests and by re-sorting the items according to the outcome of these tests (Table 2 - Table 5), we bring our sections again within limits (Table 6: 3 items in section 1, 4 items in section 2, and 4 items in section 3). This will be illustrated later as screen charts in Fig. 1 and Fig. 2.

Table 1. Inserting the new word "arrangement"

A	B	C
English	**Japanese**	**Section**
avoid	sakeru	1
minimum	saishou no	1
efficient	koukateki na	1
predecessor	zenninsha	2
linguist	gengogakusha	2
percentage	paasento	2
resemble	niteiru	2
retain	hoji suru	2
decrease	heru	2
capacity	youryou	3
arrangement	haichi	1, if space available

Assume we want to learn the new word "arrangement" in Japanese. We insert this word "arrangement", its Japanese translation "haichi" as well as its initial section number "1" at the end of our existing learning file. And sort (by Excel).

Table 2. First self-test of four items from section 1…

A	B	C
English	**Japanese**	**Section**
avoid	*sakeru*	1
minimum	*saishou no*	1
efficient	*koukateki na*	1
arrangement	*haichi*	1
predecessor	zenninsha	2
linguist	gengogakusha	2
percentage	paasento	2
resemble	niteiru	2
retain	hoji suru	2
decrease	heru	2
capacity	youryou	3

However, with this new item "arrangement", section 1 contains 4 items: Overflow! A first self-test of these four (shown in italics) will help to thin it out. If we remember a word's translation, we insert "2", if we do not, we leave it at "1". Let us try …

Table 3. Result of first self-test: 50% hit rate

A	B	C
English	**Japanese**	**Section**
avoid	*sakeru*	2
minimum	*saishou no*	2
efficient	*koukateki na*	1
arrangement	*haichi*	1
predecessor	zenninsha	2
linguist	gengogakusha	2
percentage	paasento	2
resemble	niteiru	2
retain	hoji suru	2

decrease	heru	2
capacity	youryou	3

Good job! We knew "avoid" and "minimum". The two were therefore be given the next higher section number 2 (by us, manually). The remaining items "efficient" and "arrangement" stay in section 1. Then sort again.

Table 4. Second self-test of four items from section 2…

A	B	C
English	**Japanese**	**Section**
efficient	koukateki na	1
arrangement	haichi	1
avoid	sakeru	2
minimum	saishou no	2
predecessor	*zenninsha*	2
linguist	*gengogakusha*	2
percentage	*paasento*	2
resemble	*niteiru*	2
retain	hoji suru	2
decrease	heru	2
capacity	youryou	3

Section 1 is now within limits, however, section 2 (8 items, 2 above the limit) became too full. A second self-test of some of its items (e.g. the four shown in italics) will help. If we know a word, we insert "3", if we do not know it, we insert "1". Let us try …

Table 5. Results of second self-test: 75% hit rate

A	B	C
efficient	koukateki na	1
arrangement	haichi	1
avoid	sakeru	2
minimum	saishou no	2
predecessor	*zenninsha*	3
linguist	*gengogakusha*	1
percentage	*paasento*	3
resemble	*niteiru*	3
retain	hoji suru	2
decrease	heru	2
capacity	youryou	3

Thumbs up! We knew all except the word "linguist": Therefore we inserted "1" right of "gengogakusha" and "3" right of the other three items "zenninsha", "paasento" and "niteiru". And sort again.

Table 6. Insertion and repetition cycle completed!

A	B	C
English	**Japanese**	**Section**
efficient	koukateki na	1
arrangement	haichi	1
linguist	gengogakusha	1
avoid	sakeru	2
minimum	saishou no	2
retain	hoji suru	2
decrease	heru	2
predecessor	zenninsha	3
percentage	paasento	3
resemble	niteiru	3
capacity	youryou	3

Done! All sections are again within their limits. The items "efficient", "arrangement" and "linguist", which we did not know in our self-tests, ended up in section 1. These three items will be repeated next, since the next new word inserted will cause section 1 to overflow again.

In this example, the insertion of one new item resulted in the repetition of 8 items (4 of section 1 in 1st self-test plus 4 of section 2 in 2nd self-test) as well as in the concentration of "forgotten" words in section 1. This is reasonable, because as soon as we insert the next new word, we will do a new self-test of section 1 and repeat these critical items again.

By this method, learning one new word is always associated with a systematic repetition of some older words (some items already in the worksheet) in a semi-automatic way: manual insertion of the new words and manual modification of section numbers of repeated words, but automatic sorting. Critical items (those which failed the self-tests) will be repeated most frequently. (Italics and bold are only used for illustrative purposes. These formatting features are not necessary in the learning file.)

Tutorial part

Here you can find the step-by-step description of

- how to create the learning file,
- how to insert a new item and sort, and
- how to add some statistics.

For more basic explanations related to Excel, please refer to any reference manual for that software, even if it is an older version (Nelson & Weverka, 1997). A Microsoft Office Excel (2003) for Windows XP was used. (If you use Mac OS instead of Windows, there might sometimes be some minor differences, written in parenthesis. On the Mac, I used Microsoft (R) Excel X for Mac (R) ©2002.)

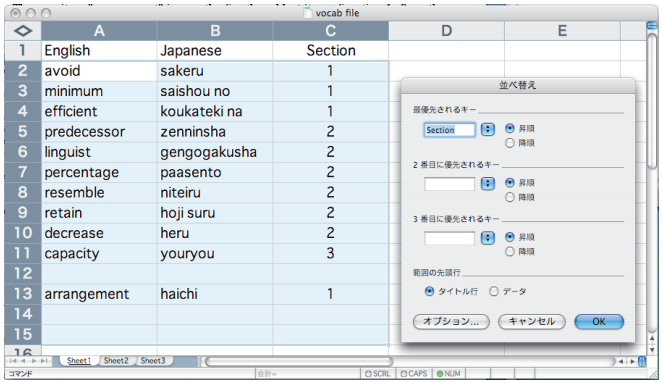
How to create the learning file

1. Start Excel. A new *workbook* window will pop up. (By default, the workbook has three sheets. At the bottom of the screen you can see three *Sheet tabs*. The first one is highlighted which means you are now looking at *Sheet 1*.)
2. Select a range which is three columns wide (A, B, C) and 15 lines high (1 … 15). (Push the mouse button while in cell A1, drag the mouse to cell C15 and release the mouse button. You can see the *selected range* in a different color, except cell A1, which is the *active cell.*)
3. Choose "Selection" in the Zoom *drop-down list box* of the *Standard Toolbar.*
(This will enlarge the selected range to maximum size within your window.)
4. Start to enter the data from Table 1 line by line; use the *Tab key* (->) after each word or number. (Type "English", ->, "Japanese", ->, "Section", ->, "avoid", ->, and so forth. As long as the 3 x 15-cell field is selected, pressing the Tab key will activate consecutive cells from the selected range: this is very convenient for entering data. In the case of typing errors, the key combination Shift-Tab will cause the previous cell to be activated. If the words do not fit completely into the cells, change either the font size or the column width.)
5. Save the workbook.

How to insert a new item and sort

If you copied Table 1 completely, the new item "arrangement " is already there (Please enter just "1" instead of "1, if space available" in the last line). We just need to sort according to the section number:

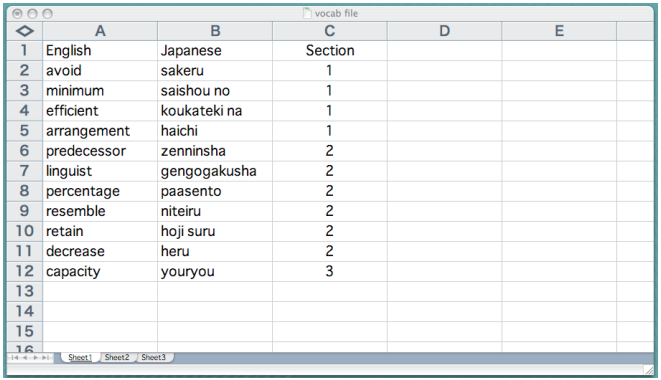
1. Select columns A, B and C.
2. Choose the *Sort* command (in Japanese: “並べ替え”) from the *Data* menu (in Japanese: “データ”). (A pop-up window will emerge. See Fig. 1.)



	A	B	C
1	English	Japanese	Section
2	avoid	sakeru	1
3	minimum	saishou no	1
4	efficient	koukateki na	1
5	predecessor	zenninsha	2
6	linguist	gengogakusha	2
7	percentage	paasento	2
8	resemble	niteiru	2
9	retain	hoji suru	2
10	decrease	heru	2
11	capacity	youryou	3
12			
13	arrangement	haichi	1
14			
15			

Figure 1. Screen shot before sorting (corresponds to Table 1)

3. In the upper box, the Sort-by *drop-down list box*, choose “Section”.
(We want to order our data according to the section numbers in ascending order, in Japanese: “昇順”. Make sure, that close to the bottom “My list has - Header row” is selected, in Japanese: “タイトル行”.) Then click “OK”.
4. The new item “arrangement” is now the fourth and last item of section 1 (See Fig. 2). Save the workbook.



	A	B	C	D	E
1	English	Japanese	Section		
2	avoid	sakeru	1		
3	minimum	saishou no	1		
4	efficient	koukateki na	1		
5	arrangement	haichi	1		
6	predecessor	zenninsha	2		
7	linguist	gengogakusha	2		
8	percentage	paasento	2		
9	resemble	niteiru	2		
10	retain	hoji suru	2		
11	decrease	heru	2		
12	capacity	youryou	3		
13					
14					
15					
16					

Figure 2. Screen shot after sorting (corresponds to Table 2)

How to add statistics

The capacities of sections 1, 2 and 3 are 3, 6 and 15 items, respectively. Section 1 contains 4 items, it overflows. We would like Excel to count the number of items in section 1 and all other sections automatically. We can use a function that counts how often the number "1" appears in column C - this is just the number of items contained in section 1. The

simplest method is to use column D and E for that purpose as follows (See Table 7):

1. Enter "Section", "1", "2" and "3" into the cells D1, D2, D3 and D4.
2. Enter "Items" into cell E1.
3. Select cell E2.
4. Click on the *Paste Function* icon (fx) in the *Formula Bar* (Mac: Standard Toolbar).
5. In the pop-up window, select *Statistical* in the upper part (Mac: on the left side), which defines the *Function category* and then select *Countif* in the lower part (Mac: on the right side), which is the *Function name*. Then click "OK".
   (This function counts the number of cells within a range that meet a given condition, in our case, that contain the number "1".)
6. A new window pops up. Either enter "C:C" in the upper *Range* box (defines the whole column C as the range), or click into the column header of column C in Sheet 1 (same effect). Now – by clicking the Tab key – proceed to the lower *Criteria box* and then click into cell D2 in Sheet 1 ("D2" will appear in this box. A so-called *relative reference* is being used). Then click "OK".
7. The number "4" should now appear in cell E2, which is the number of items in section 1. Copy the content of cell E2 and paste it into cells E3 and E4.
8. You should now see the correct results "6" and "1" in cells E3 and E4 as shown in Table 7 (the formulas in E2, E3, E4 should now be "=COUNTIF(C:C,D2)", "=COUNTIF(C:C,D3)", "=COUNTIF(C:C,D4)").
   Hint: A double click in a cell with a formula will make that formula visible. What you usually see in a cell containing a formula is just the result of the applied formula.

Table 7. Statistics of the learning file

	Vocabulary items in three sections			Statistics	
	A	B	C	D	E
1	English	Japanese	Section	Section	Items
2	avoid	sakeru	1	1	4
3	minimum	saishou no	1	2	6
4	efficient	koukateki na	1	3	1
5	arrangement	haichi	1		
6	predecessor	zenninsha	2		
7	linguist	gengogakusha	2		
8	percentage	paasento	2		
9	resemble	niteiru	2		
10	retain	hoji suru	2		
11	decrease	heru	2		
12	capacity	youryou	3		

By using the function "COUNTIF" in column E, the number of items in each of the three sections can be calculated automatically: Modifying any section number in column C will instantaneously cause the numbers in column E to change accordingly.

Effective tips for using your Excel worksheet

1. During the self-test: Set the zoom-factor to a high value such that you can only see one cell at a time on the screen (as with flashcards, where you can only see one side at a time). Another option is to decrease the window size, such that only one cell is visible. A third option is to add an extra wide column between the English word and the Japanese word. This extra column also hides the Japanese words and gives the opportunity for *writing* the word to be recalled (proposal of a workshop participant's sister).

2. You can also add color to the statistics: If the capacities of the sections are contained in the worksheet (e. g. in column F), *Conditional Formatting* in the *Format* menu can be used in such a way that cells turn red whenever the related sections overflow.

3. As more and more items are entered into the file, and as the number of cells used for statistics increases, it might be better to use an extra worksheet - *Sheet 2* - for the statistics. (In fact, the workshop version of the learning file contained data and statistics separately in two worksheets. A workshop participant proposed the simpler version explained in this paper.)

4. Leitner suggests inserting a new or repeated card at the back of a section; this guarantees that all other cards of the section will be repeated before that card. In our Excel worksheet, such strict temporal ordering is not yet realized. It can be realized by inserting the date additionally, when an new item is entered, or by updating it, when this item is being repeated. (This has to be done manually, in addition to changing the section number.) Table 8 shows such an extended worksheet with an additional date column D after section column C. (The statistics can be moved to E and F.) The date can easily be inserted in a cell by pressing the CTRL key and colon (:) or semicolon (;) simultaneously (Mac: CMD key and minus (-)).
Now, four columns A to D have to be sorted out. When sorting, two sorting criteria have to be used, the section number "Section" as first criterion (yields all lines belonging to the same section numbers being grouped together), and "Date" as second criterion (yields the lines belonging to a given section being ordered by ascending date, the "oldest" line at the top).
Table 9 shows the result of sorting, being slightly different from the sorting result in Table 4.
It is up to the user to decide, whether this advantage in ordering of lines is worth the additional burden of having to update manually two cells when repeating one item.

Table 8. Adding the current date to each repeated item

A	B	C	D
English	**Japanese**	**Section**	**Date**
avoid	*sakeru*	2	3/15/2006
minimum	*saishou no*	2	3/15/2006
efficient	*koukateki na*	1	3/15/2006
arrangement	*haichi*	1	3/15/2006
predecessor	zenninsha	2	3/13/2006
linguist	gengogakusha	2	3/13/2006
percentage	paasento	2	3/13/2006
resemble	niteiru	2	3/14/2006
retain	hoji suru	2	3/14/2006
decrease	heru	2	3/14/2006
capacity	youryou	3	3/12/2006

Table 8 corresponds to Table 3. In addition to changing the section numbers of the four repeated items (in italics), the date values have to be updated manually (entering "today's" date, 3/15/2006, by using the key combinations described in the text). Then sort again, using "Section" as first sorting criteria, "Date" as second, both ascending.

Table 9. Results of testing using dates as second sorting criteria

A	B	C	D
English	**Japanese**	**Section**	**Date**
efficient	koukateki na	1	3/15/2006
arrangement	haichi	1	3/15/2006
predecessor	*zenninsha*	2	3/13/2006
linguist	*gengogakusha*	2	3/13/2006
percentage	*paasento*	2	3/13/2006
resemble	*niteiru*	2	3/14/2006
retain	hoji suru	2	3/14/2006
decrease	heru	2	3/14/2006
avoid	sakeru	2	3/15/2006
minimum	saishou no	2	3/15/2006
capacity	youryou	3	3/12/2006

Table 9 corresponds to Table 4, however, since the date was used as second sorting criteria, the topmost items of section 2 are automatically the older ones of section 2, as in Leitner's card file box, in which the cards in the front of every section are the oldest ones of that section.

Other possible modifications are adding phrases, small sentences or hints like synonyms, antonyms, homonyms, etc. in additional columns. Furthermore, such columns could contain example sentences or semantic categories. Categories could help learners to use the learning file from time to time in an alternate mode, e. g. for the preparation of given situations (e. g. the categories "job application" or "rendezvous"), by repeating only the items of a given category.

A word of caution

A high hit rate during repetition – which is typical when starting to use this file – is encouraging, since the learner can virtually "feel" her or his knowledge growing as she or he is moving many cards back to higher section numbers. Furthermore, as long as there are not many items in the file, which is mainly during the first days or weeks of usage,

there is still a lot of empty space in the higher sections. This fact also tends to be motivating, since there is a good balance between new items to be inserted and old items to be repeated.

However, as the first few sections get full, it might become impossible to insert new items before repeating many old ones. Additionally, if the learning file has not been used for a while, there is a good chance for many items to return to section 1, because the learner will fail to recall them (low hit rate). This phenomenon causes additional congestion in section 1 and can discourage the learner.

A quick fix might be to introduce the new section number "0" for items, which will temporarily be taken out of the loop (e.g. items that are not high frequency words). Another general advice would be to spend about the same amount of time every day for learning with the file. However, these measures are not really a remedy of the basic problem: There is no royal road to learning.

Conclusion

This paper described how we can learn and repeat vocabulary in a systematic way with an Excel file. A repetition cycle consists of the three steps: (a) the attempt to recall the words from our memory; (b) setting a certain section number for each word according to whether we knew it or not; and (c) re-sorting the worksheet according to these section numbers. The order of words in this worksheet then reflects the depth of our knowledge of these words: The words at the very top of the worksheet represent words, which we do not know yet. We should repeat these items often. The lower the words are positioned in the worksheet, the better we know them. We still need to repeat these occasionally, but less and less frequently.

Markus Rude is currently associate professor at Tsukuba University, where he teaches German. He also taught German and English at Dokkyo and Rissho University. His interests are pronunciation and the usage of music and computers for language teaching.

References

Laufer, B., Meara P., & Nation, P. (2005). Ten best ideas for teaching vocabulary. *The Language Teacher*, 29 (7), 3-6.

Leitner, S. (1972). *So lernt man lernen*. Freiburg: Herder.

Meara, P. (2004). Modelling vocabulary loss. *Applied Linguistics,* 25 (2), 137-155.

Nelson, S. L., & Weverka, P. (1997). *Microsoft Office 97: The complete reference*. Berkeley: McGraw-Hill.

Raedle, P. (2005). *VTrain. The ultimate Vocabulary Trainer*. [Online] Available: <www.paul-raedle.de/vtrain/sci.htm>.

Rude, M. (2004). An Excel tool against the forgetting of vocabulary just learned. *Languaging! The Exploratory Learning and Teaching Newsletter of Dokkyo University,* (4), 14-18. [Online] Available: <www.geocities.com/languaging/>.

Straub, J. (1997). Gedächtnis. In Straub, J., Kempf, W., & Werbik, H. (Eds.), *Psychologie. Eine Einführung* (pp. 250-254). München: Deutscher Taschenbuchverlag (dtv).