MANIPULATING L2 LEARNERS' ONLINE DICTIONARY USE AND ITS EFFECT ON L2 WORD RETENTION

Elke Peters
Katholieke Universiteit Leuven

ABSTRACT

This study explored the effect of two enhancement techniques on L2 learners' look-up behaviour during a reading task and word retention afterwards amongst Flemish learners of German: a Vocabulary Test Announcement and Task-induced Word Relevance. Eighty-four participants were recruited for this study. They were randomly assigned to one of two groups: 1) not forewarned of an upcoming vocabulary test (incidental condition) or 2) forewarned of a vocabulary test (intentional condition). Task-induced Word Relevance was operationalized by a reading comprehension task. The relevance factor comprised two levels: plus-relevant and minus-relevant target words. Plus-relevant words needed to be looked up and used receptively in order to answer the comprehension questions. In other words, the reading comprehension task could not be accomplished without knowing the meaning of the plus-relevant words. The minus-relevant target words, on the other hand, were not linked to the reading comprehension questions. Our findings show a significant effect of Test Announcement and Word Relevance on whether a target word is looked up. In addition, Word Relevance also affects the frequency of clicks on target words. Word retention is only influenced by Task-induced Word Relevance. The effect of Word Relevance is durable.

INTRODUCTION

Computer technology and CALL offer promising possibilities for both language learning and research into language acquisition. Students can read online texts with a variety of learning aids at their disposal. They can look up the meaning of a word in an online dictionary, or they can practice vocabulary by using online materials and resources (e.g. concordance, on-line and interactive exercise, etc.) (Horst, Cobb, & Nicolae, 2005). Moreover, the advent of online dictionaries provides researchers with an unobtrusive instrument to track students' look-up behaviour, i.e., whether they look up a word and how often they do so (Hulstijn, 2000). Because many studies have shown beneficial effects of (online) dictionaries on word retention (Al-Seghayer, 2001; Chun & Plass, 1996; Hulstijn, Hollander, & Greidanus, 1996; Knight, 1994; Laufer & Hill, 2000), we wanted to investigate whether we could manipulate students' online look-up behaviour through a vocabulary test announcement on the one hand and a reading comprehension task, comprising textually explicit questions, on the other hand. In addition, we wanted to explore what the effects would be on short-term and long-term word retention. Does the announcement of a post-reading vocabulary test result in more frequent look-up behaviour? Does it result in better word retention? Are relevant words looked up more and remembered better than irrelevant words because students need them for answering reading comprehension questions?

PREVIOUS RESEARCH

It is generally acknowledged that L2[1] learners need to notice language features in order to acquire them (Schmidt, 2001). When students want to learn new words for instance, they need to notice unknown words and pay sufficient attention to them. Retention of new words is further determined by the way in which these words are processed, whereby deeper and more elaborate processing results in better word

retention (Hulstijn, 2001). The Involvement Load Hypothesis (Hulstijn & Laufer, 2001; Laufer & Hulstijn, 2001), for instance, provides a framework in which an attempt is made to measure the task-induced "involvement load" or depth of processing of incidentally acquired words. The construct of task-induced involvement consists of one motivational and two cognitive components: Need, Search, and Evaluation. Need, the motivational component, refers to the need to achieve, to carry out a task. Need can be moderate, when it is externally driven, or strong, when it is self-imposed. Both Search and Evaluation refer to the cognitive dimensions of the construct of task-induced involvement. Search is the attempt to find the meaning of a word or to find the word-form for a specific concept by consulting a dictionary. Evaluation "entails a comparison of a given word with other words, a specific meaning of a word with its other meanings" (Laufer & Hulstijn, 2001, p.14). As with need, evaluation can be moderate or strong.

Dictionary Use and Word Retention

When L2 learners notice an unknown word, one of the possible strategies is consulting a dictionary. Although not all looked-up words are remembered (Hulstijn et al., 1996), many studies have shown the beneficial effects of dictionary use. Hulstijn et al. investigated the differential effects of marginal glosses and a dictionary on incidental vocabulary acquisition with advanced L2 learners. Students were randomly assigned to one of three conditions: marginal glosses group, dictionary group, and control group. They were instructed to read a text (1,306 words, 16 target words) and to answer comprehension questions afterwards. Students were not forewarned that they would be tested on vocabulary used in the text. The findings indicated that dictionary consultation was low. In addition, retention scores were higher in the marginal glosses group than in the dictionary and control group. However, in the few cases that students in the dictionary group did consult the dictionary, their retention scores were higher than those of the marginal glosses group.

With the advent of CALL and the Internet, more research started to focus on the use of electronic dictionaries and its effects on incidental vocabulary acquisition. Knight (1994) investigated the effect of dictionary use on word retention in an incidental vocabulary learning condition. Students who used a dictionary acquired more word meanings than those who did not. Low verbal ability students in particular benefited from the provision of an online dictionary. Laufer and Hill (2000) have shown that the use of an electronic dictionary has a positive effect on incidental vocabulary learning. Because students may not use a dictionary because of the time involved or the disruption of the flow of reading, the provision of an electronic dictionary may be a good alternative to a paper dictionary. Laufer and Hill argue that if a pedagogical tool is popular with students, chances are it will also be beneficial for learning. In addition, their study showed better retention results than some other dictionary studies (Chun & Plass, 1996; Hulstijn et al., 1996; Knight, 1994), which may be attributed to the fact that students could choose between different types of dictionary information (L1 translation, L2 synonym, pronunciation). Other studies incorporated images, video, and audio in their dictionaries, yet the effects of multimedia annotations on word retention are not clear-cut. Chun and Plass investigated incidental vocabulary learning through a reading comprehension task within a CALL-environment, named *Cyberbuch*. They provided text, text and image, and text and video as learning aids. Highest retention scores were found for the "text and image" annotations. In a study conducted by Al-Seghayer (2001), "text and video" proved to be most effective for vocabulary acquisition. Like Laufer and Hill, the studies by Chun and Plass, and Al-Seghayer, stressed the benefits of multiple annotation types.

Laufer and Hill (2000) highlight the importance of the following three criteria when carrying out research on CALL-based dictionary use:

1) the task cannot be carried out without knowing the meaning of the target words. Relevance of target words will increase the chance of dictionary consultation.

2) the necessity of log files, which provide us with data of students' look-up behaviour

3) participants need to be given the opportunity to either consult a L2 synonym or definition, or a L1 translation.

They argue that studies that meet the above criteria can provide insight into the effect of dictionary use on vocabulary learning, the relationship between retention of looked-up words and the type of dictionary information selected, and between retention and the number of times a word is looked up.

Manipulating Tasks and Task Instructions

In spite of the beneficial effects, retention of word meanings may remain low, even when the word is consulted in a dictionary (Hulstijn et al., 1996). Electronic dictionaries, especially, are said to induce shallow processing of the dictionary information (Laufer & Hill, 2000; Rieder, 1998). As a result, some researchers have focused on how different tasks can be used to manipulate L2 learners' look-up behaviour and word retention. De Ridder (2002), Hulstijn (1993), and Laufer and Levitzky-Aviad (2003) investigated how different reading comprehension tasks influenced students' look-up behaviour. Hulstijn's findings, viz. that the type of task did not affect students' look-up behaviour, were not corroborated by De Ridder nor by Laufer and Levitzky-Aviad. Furthermore, De Ridder investigated the effect of visible and invisible hyperlinks on students' clicking behaviour, text comprehension, and word retention. Students were inclined to click more often on words in a text when visible hyperlinks were provided, but this more frequent clicking did not result in better text comprehension nor in better word retention.

Rieder (1998) conducted an experiment in which students had to focus either on text comprehension or on new vocabulary. Students who had to attend to vocabulary looked up more words. However, this did not affect word retention, but it did influence text comprehension in a negative way. In a small-scale study, Peters (2006a) investigated the effect of Test Announcement (presence or absence) on L2 learners' online look-up behaviour, word retention, and text comprehension. Before students had to read a text on a computer screen, they were assigned to one of four subject groups: 1) forewarned of a vocabulary test, 2) forewarned of a comprehension test, 3) forewarned of both a vocabulary and a comprehension test, and 4) not forewarned of any test. Hence, in this study the presence or absence of a test announcement created situations for incidental and intentional learning. The labels *incidental* and *intentional* were used strictly methodologically, as defined by Hulstijn (2001, 2003). *Incidental* means that participants are not forewarned of an upcoming test, whereas the label *intentional* refers to the fact that participants are explicitly forewarned of an upcoming test. Peters found that Test Announcement affected neither word retention nor text comprehension. Test Announcement did not have an effect on students' dictionary use when four groups were compared. However, when only two groups were compared (i.e. those forewarned and those not forewarned of a vocabulary test), students in the intentional vocabulary learning condition (= group 1 and 3) looked up more words than students in the incidental vocabulary learning condition (= group 2 and 4).

Hulstijn (1993) and Laufer and Levitzky-Aviad (2003) investigated the effect of Word Relevance on students' look-up behaviour. They defined a relevant word as a word whose meaning needed to be understood to answer a reading comprehension question, whereas an irrelevant word was not linked to a comprehension question. The findings showed that L2 learners looked up more words when these words were deemed relevant for task completion than when these words were not considered relevant.

RESEARCH QUESTIONS

The research to date has tended to investigate L2 learners' look-up behaviour under incidental learning conditions. Moreover, research has only been conducted on the effect of Word Relevance on L2 learners' dictionary use. So far, no study has investigated the effect of Word Relevance on word retention. This study aimed at investigating the effects of Test Announcement and Word Relevance on students' look- up behaviour and word retention.

In this study, we were guided by the following research questions with regard to students' look-up behaviour.

1) What is the effect of Vocabulary Test Announcement on students' look-up behaviour?
2) What is the effect of Word Relevance on students' look-up behaviour?
3) Is there an interaction effect between Test Announcement and Word Relevance on students' look-up behaviour?

We expect to find that our results will corroborate Peters' (2006a, 2006b) findings, viz. that students who are forewarned of a vocabulary test will look up more words than students who are not forewarned. In addition, words deemed relevant will be looked up more than words not deemed relevant, as demonstrated by Hulstijn (1993) and Laufer and Levitzky-Aviad (2003). The third research question addresses the interaction between Test Announcement and Word Relevance. We expect to find that students in the intentional vocabulary learning condition will look up more minus-relevant words than students in the incidental vocabulary learning condition.

The following research questions with regard to word retention were also addressed in this study.

4) What is the effect of Vocabulary Test Announcement on short-term and long-term word retention?
5) What is the effect of Word Relevance on short-term and long-term word retention?
6) Is there an interaction effect between Test Announcement and Word Relevance on word retention?

With regard to word retention, we hypothesize that students in the intentional condition will perform better than students in the incidental condition on both immediate and delayed vocabulary posttests. Short-term word retention was measured in immediate posttests, whereas long-term word retention in two delayed vocabulary tests, viz. one week and two weeks after the immediate ones. A further prediction concerns the relevance factor; words deemed relevant will be remembered better than words not deemed relevant. We expect that this effect will be durable, as measured in the two delayed vocabulary tests. The sixth research question centres around the interaction between Test Announcement and Word Relevance. We expect that students in the intentional condition will acquire more minus-relevant words than students in the incidental condition.

In addition, the following two research questions were also addressed.

7) Do students remember words after having looked them up in an online dictionary during a reading comprehension task?
8) Is there a relationship between students' frequency of lookups and word retention?

We hypothesize that dictionary consultation will affect word retention in a positive way. In addition, our prediction is that the higher the frequency of lookups, the higher the word retention scores will be.

RESEARCH METHODOLOGY

Participants

The participants in this study consisted of 84 students of German (58 females and 26 males) from two different Belgian universities and three Belgian university colleges. The university students studied German linguistics and literature, whereas the college students studied Applied Language Studies (= German translation studies). All students were considered high-intermediate students because they were recruited from the second, third, or last year of study. The participants' first language was Dutch.

Design

An experimental design with repeated measures was adopted, in which we focused on two enhancement techniques: Test Announcement (presence or absence) and Task-induced Word Relevance (plus-relevant or minus-relevant) (see Table 1).

Table 1. Research Design

	Test Announcement (between-subject variable)	**Word Relevance (within-subject variable)**
1	Incidental condition (INCID) Not forewarned of a test	Plus-relevant target words Minus-relevant target words
2	Intentional condition (INTENT) Forewarned of a test	Plus-relevant target words Minus-relevant target words

In order to operationalize Test Announcement, a between-subject variable, we distinguished between two subject groups. The first group constituted the incidental vocabulary learning condition and was not forewarned of a vocabulary test. This group was only told that they had to carry out a text comprehension task while reading a text. The second group, on the other hand, formed the intentional vocabulary learning condition. In contrast to the first group, this group was forewarned that after having answered the comprehension questions with the text at their disposal, they would be tested on the vocabulary used in the text (see also Previous Research section for definitions of *incidental* and *intentional*).

The second independent variable, Word Relevance, constituted a within-subject factor comprising two levels, viz. plus-relevant and minus-relevant target words. Task-induced Word Relevance was operationalized by having learners answer comprehension questions, for which they needed to know the meaning of half of the target words. In other words, students had to consult the meaning of eight plus-relevant words in an online dictionary in order to be able to answer the reading comprehension questions. The minus-relevant target words were not related to the comprehension questions. Hence, the relevance factor was operationalized by the reading comprehension questions, directing students' attention to eight of the sixteen target words, viz. the plus-relevant target words. The reading comprehension task was composed in such a way that it could not be carried out without knowing the meaning of the plus-relevant target words.

Since we were also interested in whether Test Announcement and Word Relevance would have a durable effect on word retention, a third independent variable came into play: Time. Time consisted of three levels: immediate word retention, word retention after one week, and word retention after two weeks. This allowed us to investigate how word retention would change over time. This within-subject variable, Time, was not taken into account in the analyses of the students' look-up behaviour. It only applied to word learning.

Students' look-up behaviour and word retention constituted the dependent variables.

Materials

The experiment was conducted on computer. A computer programme was specifically developed for this purpose using the software *QuestionMark Perception*. A Java-application was written by a computer scientist and added to the software in order to keep log files of students' lookups. The computer programme was pretested in a pilot study (Peters, 2006a). We selected the text *Ossis sind Türken* (Staud, 2003) (East-Germans are Turks), an article from the German newspaper *Die Zeit*. The text, which was shortened from 1,556 to 1,096 words, was only available on a computer screen. Students could look up unknown words[2] in an online researcher-developed dictionary by clicking on them with the mouse. When clicking on a word, a pop-up window with a German definition and a Dutch translation would appear (see

Figure 1). The words that could be looked up did not have visible hyperlinks in order to prevent students from clicking too frequently (see also De Ridder, 2002).

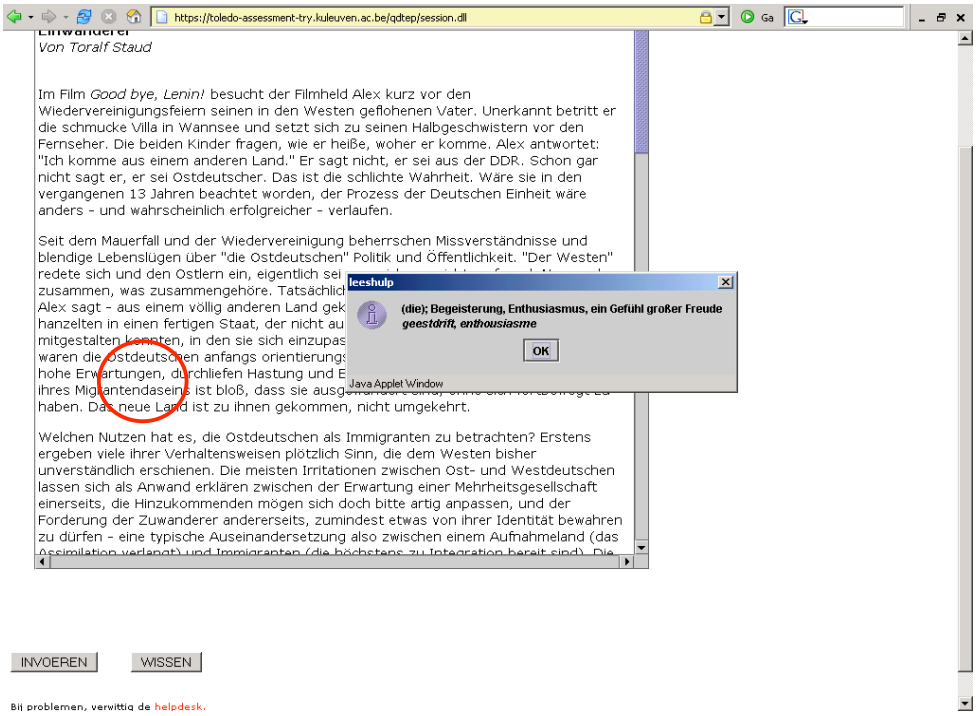

Figure 1. Screenshot of text and pop-up window of pseudoword *Hastung*

The 16 target words, split up in eight plus-relevant and eight minus-relevant words, were all replaced by pseudowords in order to ensure that no student knew the target words. German morphology was respected when creating the pseudowords so that they looked like normal, existing German words to our participants. We expected that none of the students would notice that the target words were German pseudowords because the text contained several low-frequency words. The meaning of the 16 target words could hardly be inferred from the context. As a consequence, students had to consult the dictionary if they wanted to know the meaning of a target word.

Text excerpt with plus-relevant target word in bold:
Nach 1989 gab es für die Ostdeutschen fast nichts, was sich nicht änderte: ein neues politisches System hielt Einzug. Aus den Regalen der Kaufhallen verschwanden die gewohnten ***Kölche****, aus dem Radio die gewohnten Sender.*

Comprehension question related to this excerpt:
Wat verdween er na 1989 uit de rekken van de winkels? (*What disappeared from the supermarkets after 1989?)*

Answer:
De gewone voedingsmiddelen (*the normal food*)
"Food" (*Kölche*) is one of the eight plus-relevant target words.

Figure 2. Example of one of the reading comprehension questions

A reading comprehension task was developed to operationalize the relevance factor. Because this task contained only seven textually explicit questions[3], the text remained available for the students while they

answered the comprehension questions (see Figure 2). For one question, two plus-relevant target words needed to be looked up. The questions (in Dutch) had to be answered in students' L1, Dutch, to ensure that students would consult the online dictionary.

Students scored generally well on the reading comprehension task (mean = 5.57, with a range from 2.5 to 7 [the maximum score]; *SD* = 1.06). When students did not look up a plus-relevant word for answering the accompanying reading comprehension question, they never provided a correct answer, which contained the Dutch translation of the target word, except for one student[4]. In short, the design and materials of this study meet the three criteria postulated by Laufer & Hill (2000) (see Previous Research section).

Data Collection Instruments

Vocabulary size test: It is known that a L2 learner's vocabulary size is important when acquiring (new) vocabulary (Horst, Cobb, & Meara, 1998). Therefore, students' vocabulary size was measured with a self-designed test based on the German frequency list of the Institut für Deutsche Sprache - Mannheim to account for possible proficiency differences. We developed a test containing 50 multiple choice items presented in isolation. For every test item, we provided four options: the correct solution and three distractors. The test covered the frequency bands from 4,001 to 10,000 words. We chose not to focus on the first 4,000 words because all participants had studied German for some years. Students' total score (max. = 50) on the vocabulary size test would function as covariate in the statistical analyses.

Vocabulary retention tests: Vocabulary retention was measured in four receptive vocabulary tests (Figure 3) in order to capture partial vocabulary retention because after one exposure, L2 learners cannot be expected to fully know the meaning of the target words (Henriksen, 1999; Nation, 2001).

- LexNotTest:
  verzettern ☐ yes ☐ no
- IsolTest:
  verzettern: _____
- ContTest:
  Immigranten neigen dazu, sich **verzettert**, ausgebeutet und [...] zu fühlen.
  verzettern: _____
- MatchTest:

	a. in derselben Form bestehen bleiben *(blijven) bestaan, in stand blijven*
7. hanzeln	b. jemandem Schwierigkeiten machen *pesten, treiteren*
8. verzettern	c. zufällig in eine unangenehme Situation kommen; geraten *geraken, terechtkomen, belanden*
9. tantieren	d. in jemandem Ekel oder Widerwillen hervorrufen; abstoßen; nicht gefallen *afstoten, wegstoten*
10. verlämpfen	e. etwas wird weniger intensiv, abnehmen *afnemen, verminderen, verzwakken*
	f. eine Sache in die gewünschte Form bringen, ihr die gewünschten Merkmale geben *meevormen, mee gestalte geven*

Figure 3. Plus-relevant target word *verzettern*, tested in four vocabulary posttests

We administered first a recognition[5] test, two recall tests, and finally a second recognition test. If a word was noticed and/or looked up, students would probably remember having seen that word in the text.

Consequently, we assumed that they would be able to recognize the word form. Thus, we developed a recognition test, which offered a list of 40 word forms, each accompanied by the question whether the word had appeared in the text or not. This test was labelled "lexical noticing test" (LexNotTest). In the first recall test, we offered the target words in isolation; students were required to supply their meaning by writing a L1 translation or L2 synonym (IsolTest). In the second recall test, the target words were offered in the sentences as they had occurred in the text (ContTest). Finally, the fourth test was a matching test, in which students had to tick off the correct meaning of the target words (MatchTest). We assumed that the IsolTest would be the most difficult test and the recognition tests the easiest ones. The IsolTest, ContTest, and MatchTest were also used to test long-term word retention. They were administered one week and two weeks after the first session (see also the Procedure subsection).

In addition to the collection of quantitative data, we also collected qualitative data. Students filled in a questionnaire with retrospective questions with which we tried to elicit students' strategy use and how they had processed the dictionary information. Furthermore, seven students provided us with think-aloud protocols. These qualitative data were used to help us interpret our findings.

Data Analysis

Students' look-up behaviour was coded in two ways. First, we wanted to know whether the target words were looked up (first coding). Second, we wanted to know how many times these words were looked up (second coding), i.e. subsequent lookups were also counted. The immediate and the delayed vocabulary tests were all scored dichotomously. A correct response received 1 point, an incorrect one 0 points.

As statistical analyses, MANCOVAs with Test Announcement as between-subject and Word Relevance as within-subject variables were conducted in order to investigate the effect of either independent variable and the interaction between the two independent variables on students' look-up behaviour and word retention as measured in the LexNotTest. Vocabulary size constituted the covariate. For the IsolTests, ContTests, and MatchTests, we employed MANCOVAs with repeated measures on Test Announcement and Word Relevance to investigate how word retention had changed over the three test sessions. In the MANCOVA-tables, we refer to Test Announcement as *Group*, to Word Relevance as *Relevance*, and to the covariate Vocabulary Size as *Vocsize*. Pearson correlation coefficients were computed to determine the relationship between students' look-up behaviour and word retention. In all statistical analyses, a *p*-level of .05 was taken as the level of statistical significance.

Procedure

Data were collected in three sessions. Before starting with the experiment, participants were shown how to use the online dictionary, but they were not told that their look-up behaviour would be registered. Students were randomly assigned to one of the two treatment groups. All students started by reading their instruction on the computer screen. The incidental vocabulary learning group received the instruction that they had to read a text on a computer screen, and that they had to do a comprehension task. Students were also told that having read the text they would have to answer some questions about how they had experienced the experiment. The intentional vocabulary learning group received the same instruction, but in contrast to the first group they were forewarned of a receptive vocabulary test. Next, students started reading the text with the reading comprehension questions at their disposal. No time limit was set for the comprehension task.

The two groups took the four vocabulary tests immediately after having accomplished the reading comprehension task. Students started with the LexNotTest, followed by the IsolTest, then the ContTest, and finally the MatchTest. These tests were administered one by one. Having completed them, students had to answer retrospective questions about the experiment. Finally, they took the vocabulary size test. Except for the reading comprehension task, all tests were administered on computer.

We also tested long-term word retention. Two delayed paper and pencil tests, consisting of the IsolTest, ContTest, and MatchTest, were administered, the first time one week and the second time two weeks after the first session. Except for three students who participated on a voluntary basis, the experiment was organized during students' normal class hours.

RESULTS

Vocabulary Size Test

The average score on the vocabulary size test was 26 out of 50 (*SD* = 8). The INTENT group performed only slightly better than the INCID group (26 vs. 25.5). The vocabulary size test proved to be reliable (*N* = 83; Cronbach's alpha = .82). A two-sample *t*-test did not reveal a significant difference in vocabulary size between the two groups ($t = 0.35$; $DF = 81$; $p = .73$). We decided to retain the vocabulary size test as covariate because it correlated with students' year of study. In this way, we could control for proficiency differences. Moreover, students' vocabulary size correlated with their word retention scores as measured in the posttests.

Look-up Behaviour

As explained in the Data Analysis subsection, students' look-up behaviour of the target words was coded in two ways. First, we present students' look-up behaviour according to the first coding (looked up or not), next the results of students' look-up behaviour according to the second coding (frequency of lookups).

With regard to the first coding, students in the intentional condition (INTENT) consulted more target words in the online dictionary than students in the incidental condition (INCID) as can be seen from Table 2. The results of the MANCOVA (Table 3) showed a significant effect of Vocabulary Test Announcement ($p = .03$). However, its effect size was not very large ($\eta_p^2 = .06$).

Table 2. Words Looked up per Condition and per Word Type (First Coding) and MANCOVA

		All target words		Plus-relevant		Minus-relevant	
	N	***M***	***SD***	***M***	***SD***	***M***	***SD***
INCID	42	11	3.1	7.5	1.0	4.0	2.6
INTENT	42	13	2.7	7.7	0.7	5.1	2.3
All students	84	12	3.0	7.6	0.9	4.5	2.5

Table 3. MANCOVA of Look-up Behaviour According to First Coding

Source	***DF***	***F* Value**	***Pr***	**Partial Eta Sq**
Within-subject				
Relevance	1	23.18	<.0001*	.23
Relevance*Group	1	4.32	.04*	.05
Relevance*Vocsize	1	2.05	.16	.03
Error (within)	80			
Between-subject				
Group	1	4.82	.03*	.06
Vocsize (covariate)	1	1.83	.18	.02
Error (between)	80			

With regard to the relevance factor, students looked up an average of three plus-relevant target words more than minus-relevant ones (Table 2). As can be seen from Table 3, the difference between the two

categories of words was statistically significant ($p < .0001$). Moreover, the partial eta squared of Word Relevance ($\eta_p^2 = .23$) was much larger than the one of Test Announcement ($\eta_p^2 = .06$).

As can be seen from Table 2, students in the INTENT group looked up more minus-relevant target words compared to students in the INCID group. The MANCOVA (Table 3) showed that the interaction between Test Announcement and Word Relevance was significant ($p < .05$). A Post-hoc Tukey Kramer analysis, which adjusts for multiple comparisons, indicated indeed that the two groups differed significantly in their lookups of the minus-relevant target words.

Having shown that both Test Announcement and Word Relevance affected students' look-up behaviour according to the first coding, we also wanted to investigate whether the same would be true when all subsequent clicks were counted (second coding). Students in the intentional condition clicked more frequently on target words than students in the incidental condition (Table 4). However, Test Announcement did not yield an effect on students' frequency of clicks on target words (Table 5).

Table 4. Lookups per Condition and per Word Type (Second Coding) and MANCOVA

		All target words		**Plus-relevant**		**Minus-relevant**	
	N	***M***	***SD***	***M***	***SD***	***M***	***SD***
INCID	42	18	8.0	14	5.7	4	3.2
INTENT	42	21	6.7	14	4.5	7	3.6
All students	84	19	7.5	14	5.1	5	3.6

Table 5. MANCOVA of Look-up Behaviour According to Second Coding

Source	***DF***	***F* Value**	***Pr***	**Partial Eta Sq**
Within-subject				
Relevance	1	22.87	<.0001*	.20
Relevance*Group	1	3.07	.08	.04
Relevance*Vocsize	1	0.00	.96	.00
Error (within)	80			
Between-subject				
Group	1	2.46	.12	.03
Vocsize (covariate)	1	0.08	.78	.00
Error (between)	80			

With regard to Word Relevance, students clicked on average three times more often on plus-relevant words in comparison to the minus-relevant target words (Table 4). The results of the MANCOVA revealed a significant effect for task-induced Word Relevance as can be gleaned from Table 5.

Contrary to our expectation, students in the intentional condition did not click more frequently on minus-relevant target words since the MANCOVA did not show a significant interaction (Table 5).

Word Retention

The results with respect to word retention are presented per test type. With respect to the IsolTest, ContTest, and MatchTest, we only provide the MANCOVAs with repeated measures on Test Announcement and Word Relevance because they corroborated the results of the inferential statistics per test session, which are presented in a footnote. We also report on the relationship between students' look-up behaviour and word retention.

Lexical noticing test

The difference in word retention between students in the intentional and incidental condition was very small as can be gleaned from Table 6. A MANCOVA did not yield an effect for Test Announcement. The fact that the former condition was forewarned of an upcoming vocabulary test did not help them to recognize more target words.

Students remembered having seen more plus-relevant words in the text than minus-relevant ones (Table 6). The MANCOVA revealed that this difference was statistically significant ($p < .0001$).

Table 6. Word Retention per Condition, per Word Type and MANCOVA on the Lexical Noticing Test

		All target words		**Plus-relevant**		**Minus-relevant**	
	N	***M***	***SD***	***M***	***SD***	***M***	***SD***
INCID	42	11.83	2.64	6.98	1.47	4.86	1.70
INTENT	42	12.38	2.58	7.00	1.23	5.38	1.89
All students	84	12.11	2.61	7	1.35	5	1.81

Source	***DF***	***F* Value**	***Pr***	**Partial Eta Sq**
Within-subject				
Relevance	1	16.69	.0001*	.17
Relevance*Group	1	1.72	.19	.02
Relevance*Vocsize	1	2.19	.14	.03
Error (within)	80			
Between-subject				
Group	1	1.11	.30	.01
Vocsize (covariate)	1	4.01	.05	.05
Error (between)	80			

Students in the intentional condition hardly differed from students in the incidental condition in their retention of the plus-relevant and minus-relevant target words. The results indicated that there was no significant interaction between the two independent variables (Table 6).

Words tested in isolation[6]

On the first and third IsolTest, it was the INCID group that performed best. The INTENT group, on the other hand, obtained the best score on the second test (Table 7). The data were subjected to a MANCOVA with repeated measures on Test Announcement and Word Relevance. Only the data of 47 participants were used in this analysis because not all students were present during the second and third test session. This analysis indicated that the two groups did not differ significantly (Table 8).

As can be seen from Table 7, students supplied more correct answers for the plus- than for the minus-relevant target words on each of the three IsolTests. The MANOVA (Table 8) revealed that Word Relevance had a significant effect on word retention when tested in isolation.

With respect to possible interactions between Test Announcement and Word Relevance, students in the INTENT group remembered more minus-relevant target words in the first and second IsolTest, whereas the INCID group did so in the third IsolTest. However, there was no significant interaction, as can be seen from Table 8.

Table 7. Word Retention per Condition and per Word Type on the Three IsolTests

		IsolTest1				IsolTest2				IsolTest3			
		All target words				All target words				All target words			
	N	*M*		*SD*		*M*		*SD*		*M*		*SD*	
INCID	42	4.57		2.74		3.89		2.17		6.84		3.14	
INTENT	42	4.55		2.77		4.15		1.80		6.21		2.41	
		Plus		Minus		Plus		Minus		Plus		Minus	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
INCID	42	3.71	2.10	0.86	1.07	3.08	1.68	0.82	1.04	5.08	1.80	1.76	1.67
INTENT	42	3.62	1.20	0.93	1.28	3.36	1.48	0.9	0.83	4.66	1.42	1.55	1.35
All students	84	3.67	2.04	0.89	1.17	3.23	1.58	0.81	0.93	4.85	1.61	1.65	1.49

Note: Total Score in Upper Panel and Scores for Plus- and Minus-relevant Words in Lower Panel. IsolTest1 is based on data of 84 participants, IsolTest2 on data of 77 participants, and IsolTest3 on data of 54 participants. IsolTest 1 was administered during the first session, IsolTest2 during the second one, and IsolTest3 during the last one. *Plus* refers to the plus-relevant target words, *minus* to the minus-relevant target words.

Table 8. MANCOVA with Repeated Measures: IsolTests

Source	*DF*	*F* Value	*Pr*	Partial Eta Sq
Time (within-subject)	1	6.55	.002*	.13
Time*Vocsize	1	0.87	.42	.02
Time*Group	1	1.07	.35	.02
Relevance (within-subject)	1	24.32	< .0001*	.35
Relevance*Vocsize	1	0.03	.87	.00
Relevance*Group	1	0.09	.77	.00
Time*Relevance	1	2.44	.09	.05
Time*Relevance*Vocsize	1	1.55	.21	.03
Time*Relevance*Group	1	.07	.94	.00
Group (between-subject)	1	0.33	.57	.01
Error	45			

As can be seen from Table 7, students in both conditions scored less well on the second IsolTest compared to the first IsolTest, yet they improved their score on the third IsolTest. Surprisingly, scores on the third IsolTest were even higher than the scores on the first IsolTest (6.84 vs. 4.57 in the incidental condition, 6.21 vs. 4.55 in the intentional condition). The MANCOVA yielded an effect of Time (see Table 8). This will be further clarified in the Discussion section.

Retention of words in context[7]

Retention scores on the first ContTest were higher than on the first IsolTest (Table 9). A paired *t*-test indicated that students could supply significantly more correct answers when the target words were offered in context ($t = 16.44$; $DF = 83$; $p < .0001$).

The intentional group obtained better scores on the three vocabulary tests in which the target words were offered in context (Table 9), yet a MANCOVA revealed that the difference was not statistically significant (Table 10).

Table 9. Word Retention per Condition and per Word Type on the Three ContTests

		ContTest1				ContTest2				ContTest3			
		All target words				All target words				All target words			
	N	*M*		*SD*		*M*		*SD*		*M*		*SD*	
INCID	42	8.24		2.67		7.21		2.72		7.84		3.00	
INTEN	42	8.69		2.26		7.71		2.52		8.07		2.36	
		Plus		Minus		Plus		Minus		Plus		Minus	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
INCID	42	6.07	1.47	2.17	1.89	5.29	1.72	1.92	1.65	5.58	1.58	2.27	1.87
INTEN	42	6.45	1.23	2.24	1.57	5.64	1.35	2.08	1.64	5.59	1.38	2.48	1.68
All students	84	6.26	1.36	2.20	1.72	5.47	1.54	2.00	1.64	5.58	1.46	2.38	1.76

Note: Total Score in Upper Panel and Scores for Plus- and Minus-relevant Words in Lower Panel. ContTest1 is based on data of 84 participants, ContTest2 on data of 77 participants, and ContTest3 on data of 55 participants. ContTest1 was administered during the first session, ConTest2 during the second one, and ContTest3 during the last one. *Plus* refers to the plus-relevant target words, *minus* to the minus-relevant target words.

Table 10. MANCOVA with Repeated Measures: ContTests

Source	*DF*	*F* Value	*Pr*	Partial Eta Sq
Time (within-subject)	1	0.25	.61	.01
Time*Vocsize	1	0.12	.73	.00
Time*Group	1	0.10	.75	.00
Relevance (within-subject)	1	36.55	< .0001*	.44
Relevance*Vocsize	1	2.48	.12	.05
Relevance*Group	1	0.10	.75	.00
Time*Relevance	1	2.43	.13	.05
Time*Relevance*Vocsize	1	0.16	.96	.00
Time*Relevance*Group	1	1.33	.26	.03
Group (between-subject)	1	0.29	.60	.00
Error	46			

In the three ContTests, retention of the plus-relevant words was much higher than retention of the minus-relevant ones as can be seen from Table 9. On the immediate ContTest, students remembered four plus-relevant words more than minus-relevant ones. On the delayed ContTests, they retained on average three plus-relevant target words more than minus-relevant ones. The difference between the two categories of words was statistically significant ($p < .0001$), as can be seen from Table 10. Relevance had a larger effect on word retention when tested in context ($\eta_p^2 = .44$) compared to when tested in isolation ($\eta_p^2 = .35$).

Though students in the intentional condition remembered more plus- and more minus-relevant target words on the immediate and on the delayed vocabulary tests than students in the incidental condition, the differences were not large. The MANCOVA (Table 10) showed that there was no significant interaction between Test Announcement and Word Relevance. Hence, we did not find evidence for our hypothesis that students in the intentional condition would remember more minus-relevant target words than students in the incidental condition.

Did the retention scores on the ContTests change over time? In contrast to retention in the IsolTests, the highest retention score was found on the *first* test. The retention scores decreased on the second test, but

improved again on the third test. Unlike the case of the IsolTest, no Time effect could be revealed as can be seen in Table 10. The MANCOVA was conducted on the data of 48[8] participants.

Word retention in the matching test[9]

The highest retention scores were found on the matching test. These scores were significantly higher than the scores on the ContTest ($t = 16.21$; $DF = 83$; $p = < .0001$).

As can be seen from Table 11, students' retention scores on the three matching tests fluctuated around 12. On the immediate and third test, students in the intentional condition could match slightly more words with their definition than students in the incidental condition. On the second MatchTest, it was the INCID group that performed best. The difference, however, was not significant at the .05-level in any of the three MatchTests (Table 12).

Table 11. Word Retention per Condition and per Word Type on the Three MatchTests

		MatchTest1				**MatchTest2**				**MatchTest3**			
		All target words				**All target words**				**All target words**			
	N	***M***		***SD***		***M***		***SD***		***M***		***SD***	
INCID	42	12.19		3.01		12.95		2.49		12.12		3.04	
INTENT	42	12.88		2.50		12.26		2.50		12.69		2.58	
		Plus		**Minus**		**Plus**		**Minus**		**Plus**		**Minus**	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
INCID	42	7.02	1.41	5.17	2.14	7.32	1.02	5.63	2.02	7.19	1.06	4.92	2.35
INTENT	42	7.17	1.10	5.71	1.97	7.18	1.10	5.08	1.97	7.28	0.92	5.41	2.11
All students	84	7.10	1.26	5.44	2.06	7.25	1.05	5.35	2.00	7.24	0.98	5.18	2.22

Note: Total Score in Upper Panel and Scores for Plus- and Minus-relevant Words in Lower Panel. MatchTest1 is based on data of 84 participants, MatchTest2 on data of 77 participants, and MatchTest3 on data of 55 participants. MatchTest1 was administered during the first session, MatchTest2 during the second one, and MatchTest3 during the last one. *Plus* refers to the plus-relevant target words, *minus* to the minus-relevant target words.

Participants could match more plus-relevant target words with their definition than minus-relevant target words as can be seen from Table 11. This was true for all three MatchTests. The differences were statistically significant ($p < .0001$) (Table 12).

Did we find evidence for our hypothesis that students in the intentional condition would remember more minus-relevant target words than students in the incidental condition? Although students in the intentional condition scored better with regard to the minus-relevant target words on the first and third matching test, the difference was not large enough to be statistically significant, as can be seen from Table 12.

As can be gleaned from Table 11, students' scores did not differ much on the three matching tests. All group means were subjected to a MANCOVA to determine whether there would be a significant time effect, yet this analysis did not reveal such an effect (Table 12). However, the analysis did show a significant interaction between Time and Test Announcement because the two groups did not manifest the same retention pattern. The highest scores in the intentional group were obtained on the first MatchTest, whereas in the incidental group, students performed best on the second MatchTest.

Table 12. MANCOVA with Repeated Measures: MatchTests

Source	*DF*	*F* Value	*Pr*	Partial Eta Sq
Time (within-subject)	1	1.19	.31	.03
Time*Vocsize	1	1.50	.23	.03
Time*Group	1	5.42	.01*	.11
Relevance (within-subject)	1	15.65	< .0001*	.25
Relevance*Vocsize	1	2.47	.12	.05
Relevance*Group	1	0.10	.76	.00
Time*Relevance	1	0.78	.46	.01
Time*Relevance*Vocsize	1	0.87	.42	.02
Time*Relevance*Group	1	1.39	.26	.03
Group (between-subject)	1	0.01	.92	.00
Error	46			

Is there a relationship between the number of lookups and word retention?

If a word was looked up, what were the chances that a student would retrieve its meaning correctly on the vocabulary tests? Retention was not the same for the plus-relevant and minus-relevant target words (Table 13).

Table 13. Percentage of Mean Retention Scores if Word is Looked up

	Plus-relevant word	Minus-relevant word
IsolTest1	49%	20%
ContTest1	82%	40%
MatchTest1	91%	73%
IsolTest2	43%	14%
ContTest2	71%	32%
MatchTest2	91%	68%
IsolTest3	64%	23%
ContTest3	72%	35%
MatchTest3	91%	66%

When tested in isolation or in context, the chance of remembering the meaning was twice as high for a plus-relevant word, which was looked up, than for a minus-relevant word. This tendency remained over time. If looked up, plus-relevant words were better remembered than minus-relevant words.

Pearson correlation coefficients were computed in order to determine the relationship between the number of lookups and word retention (Table 14).

The results for the total number of target words as well as for the plus- and minus-relevant target words separately are presented in Table 14. Low to moderate correlations were found, but only the correlation coefficients for the first and second ContTest were statistically significant. Except for the first ContTest, only low and insignificant correlations were found for the plus-relevant target words. The correlation coefficients for the minus-relevant words were higher, though still low to moderate.

Table 14. Pearson Correlation Coefficients of Number of Lookups and Word Retention

	LexNotTest	**IsolTest1**	**ContTest1**	**MatchTest1**
Target words	0.20	0.07	0.39*	0.05
Plus-relevant	0.005	0.009	0.25*	0.07
Minus-relevant	0.28*	0.14	0.37*	0.03
		IsolTest2	**ContTest2**	**MatchTest2**
Target words		0.07	0.28*	0.20
Plus-relevant		-0.05	0.11	0.16
Minus-relevant		0.23*	0.33*	0.17
		IsolTest3	**ContTest**	**MatchTest3**
Target words		-0.00	-0.06	-0.03
Plus-relevant		-.002	-0.16	0.07
Minus-relevant		0.02	0.04	-0.06

Note: Correlations indicated with an asterisk are significant at the .05-level.

DISCUSSION

Look-up Behaviour

This study has shown that it is possible to manipulate L2 learners' noticing of unknown words, measured in terms of lookups. The fact that we were successful in manipulating students' dictionary use is important because noticing is the first step in the acquisition process (Hulstijn, 2001; Schmidt, 2001). When students clicked on a word, we can be absolutely sure that students noticed that word. Students who were forewarned of an upcoming vocabulary test looked up more target words than students who did not know that a test would follow. Since these students knew that a test would follow, they probably attended more to vocabulary, which was also corroborated by students' answers to one of the retrospective questions. Consequently, they clicked more readily on a word to find out its meaning. Our findings corroborate Peters' (2006a, 2006b) and Rieder's (1998) results. However, Vocabulary Test Announcement did not affect the frequency of lookups. When all clicks were counted, individual preferences prevailed over task instruction. The qualitative data revealed that some students were very actively engaged in their dictionary use and even made vocabulary their learning aim, while others, a minority of students, confined themselves to skimming the text to find the answers to the comprehension questions. In other words, the participants approached the task in a way they found appropriate, irrespective of the pre-learning instructions.

Apart from Test Announcement, Word Relevance also had an effect on whether students looked up a target word. Furthermore, the effect of relevance was even greater since it also affected the frequency of clicks. In addition, its effect size was larger compared to the one of Test Announcement. Our results are in line with Hulstijn (1993), and Laufer and Levitzky-Aviad (2003), i.e., that words which are deemed relevant are looked up more often than words not perceived relevant. Moreover, the fact that these plus-relevant words were indeed looked up is also attributable to the reading comprehension task itself because the answers to the comprehension questions had to be given in Dutch. In other words, the reading comprehension task promoted the salience of the plus-relevant target words and directed students' attention to these words. But how can we explain that some students did not look up all plus-relevant target words? When a plus-relevant target word was not looked up, this was attributable rather to the fact that students focused on a wrong (part of the) sentence than to the fact that students were unwilling to consult the dictionary. When students did not look up a plus-relevant word, they did not answer the question related to this word correctly.

In this study, we found an interesting interaction between Test Announcement and Task-induced Word Relevance. Even if a word was not deemed relevant for task completion, Vocabulary Test Announcement had an effect on whether a minus-relevant word was looked up or not. Students in the intentional condition were more aware of unknown words, which had an effect on the minus-relevant target words in particular. In contrast to students in the incidental condition, these students were more inclined to consult the meaning of a minus-relevant target word.

Word Retention

Through forewarning students of a vocabulary test, we were able to manipulate students' noticing of words, yet noticing is not sufficient for the actual acquisition of words. The findings are at variance with our hypothesis, because Test Announcement did not induce more elaborate word processing, which is also required if word learning is really to happen. Hence, we agree with Hulstijn (2001, p. 275) that it is what students actually do with words and not the absence or presence of a test which determines the success of word retention. In short, word processing in the intentional condition was not as we anticipated.

Students approach a reading task with a priority for meaning, no matter what instruction they get. The retrospective questions and think-aloud protocols revealed that there were hardly any differences in task approach or strategy use between the two groups. These data showed a "focus on content and not a focus on individual words"-strategy. In addition, very few vocabulary strategies were revealed at all. With regard to the minus-relevant words, we agree with Hulstijn (1993) that one of the possible strategies that students can use is to "ignore the word altogether" (p. 146). By forewarning students of an upcoming vocabulary test, we were not able to manipulate the way students approach a task. Our results are in line with Peters (2006a) and Sercu, Dewachter, Peters, Kuiken, and Vedder (2006), who also found that the presence of a vocabulary test announcement did not affect word retention.

Although we were not successful in fostering word retention by forewarning students of a test, we found ample evidence for the relevance factor. Not only were the plus-relevant words looked up more often, they were also better remembered than the minus-relevant words both in the short and the long term. Compared to the studies by Hulstijn (1993), and Laufer and Levitzky-Aviad (2003), which focused solely on students' look-up behaviour, our study provides additional evidence of the effect of Task-induced Word Relevance.

Although the main aim of this study was not to find empirical evidence for the Involvement Load Hypothesis (Hulstijn & Laufer, 2001; Laufer & Hulstijn, 2001; see also Previous Research section), it may shed more light on why the plus-relevant words were better remembered than the minus-relevant words. The comprehension task made the plus-relevant target words salient for learning by directing students' attention to these words. The reading comprehension questions induced a need to know the meaning of the plus-relevant target words. The value of this need was moderate because it was externally driven. The minus-relevant target words were not linked to the reading comprehension questions so they did not induce an externally driven need to know the meaning. Students consulted the online dictionary in an attempt to find the meaning of the plus-relevant target words. Moreover, log files revealed that students clicked more frequently on the plus-relevant target words. Though no different values of search are discussed in any paper, search takes on a higher value when the word is looked up more often (Laufer, personal communication, 2005). So the value of search was higher for the plus- than for the minus-relevant target words. Since all dictionary explanations were context-bound, evaluation did not play a role in our study. The higher task-induced involvement load can shed light on why students learned more plus-relevant than minus-relevant target words.

However, the Involvement Load Hypothesis does not explain the full picture. This is best illustrated with the retention chances of a word if looked up (see Table 14). We hypothesize that, irrespective of the frequency of lookups, students processed the plus-relevant target words more elaborately than the minus-

relevant target ones because students really had to do something with them. Though there is no unequivocal interpretation of "levels of processing," the richer or more elaborately a word is processed, the greater the chance that it will be remembered. This is exactly what happened with the plus-relevant target words. As a consequence, this deeper and more elaborate processing left a stronger memory trace for the plus-relevant target words. "What is critical to retention is … the richness with which the material is encoded … processing new lexical information more elaborately … will lead to higher retention than by processing new lexical information less elaborately" (Hulstijn, 2001, p. 270).

Several arguments support this assumption that the plus-relevant target words were processed more elaborately. First, apart from looking up the word, students also had to write down the translation of the plus-relevant target word in their answer to the comprehension questions. The think-aloud protocols revealed that students engaged in quite some translating when accomplishing the reading comprehension task. Laufer and Girsai (2005) conducted a study in which they investigated the effect of a translation task on incidental vocabulary acquisition, compared to message- and form-focused tasks. The translation task always yielded the highest retention scores on immediate as well as delayed posttests. The beneficial effects of Word Relevance may also be attributed to the use of translation in the comprehension task.

A second argument is that the plus-relevant target words were processed more frequently than the minus-relevant words. Students read the target word in the text (first encounter), looked it up (second encounter), reread the sentence with the plus-relevant target word (third encounter), and finally wrote down the answer (fourth encounter). The plus-relevant target words should have been processed at least four times, which was also confirmed by the think-aloud protocols. These protocols revealed that many students reread certain paragraphs to verify their answers to the comprehension questions, which may have added even more “encounters” with the plus-relevant target words.

Finally, the think-aloud protocols also revealed that students could recall their learning behaviour more easily when they were confronted with the plus-relevant target words. It is episodic memory (Baddeley, 1997) which plays a role in triggering the reactivation of a learning experience. When students were re-exposed to the plus-relevant target words, especially when these words were offered in the sentences as they had occurred in the text, students' learning behaviour (answering comprehension questions and looking up words) was reactivated and as a result, also the form-meaning connection. These arguments demonstrate that, irrespective of the number of lookups, students processed the plus-relevant words more elaborately. As a consequence, they left a stronger memory trace than the minus-relevant target words.

A final issue we want to address is how students' word retention changed over time. Performance on the three vocabulary posttests (immediate, one week, and two weeks later) indicated that word retention did not change in the same way when the words were tested in isolation or in context. In the case of the former, students obtained the highest scores on the third test. The effect of time, i.e. the increase on the last test, was statistically significant. Why was there such an improvement in the last IsolTest? The effect of retesting may have influenced the scores on the delayed tests. Several studies (see Baddeley, 1997, for survey) have shown that a retrieval practice effect exists. Each successful retrieval increases the chance that a word will be retained because each retrieval strengthens the retrieval route and the form-meaning connection (Baddeley, 1997, chapter 7). Watanabe (1997) also found an increase in retention scores on delayed vocabulary tests. Moreover, the think-aloud protocols revealed that students' test taking strategy during the first MatchTest influenced the scores on the delayed tests. Students tended to remember the contexts of the ContTest when taking the MatchTest. First, students matched the words with the definition of which they were absolutely sure. Next, they focused on the target words for which they had not yet supplied an answer. Students picked one of the remaining definitions and checked whether it fitted the context they could remember from the previous test. The fact that students could actually see the correct solution in the five options of the MatchTest contributed even more to the retesting effect. Unexpectedly, the MatchTest did not only test the target words but it also created opportunities to acquire them since some students finally established the form-meaning connections.

Relationship between students' look-up behaviour and word retention

If a plus-relevant word was looked up, retention was very high on the immediate vocabulary tests. In addition, if looked up, retention of a plus-relevant word was twice as high as retention of a minus-relevant target word. This result supports the claim that provision of a dictionary is beneficial for word learning (Al-Seghayer, 2001; Chun & Plass, 1996; Hill & Laufer, 2003; Hulstijn, 1993; Knight, 1994; Laufer & Hill, 2000). However, the frequency of clicks did not correlate very well with word retention, which is in line with Chun and Plass (1996) and Laufer and Hill (2000). The lack of high and significant correlations could be attributed to the use of an electronic dictionary. Although electronic dictionaries do not disrupt the flow of reading, they might induce shallow processing and detrimental word retention (Laufer & Hill, 2000, p. 67; Rieder, 1998). Because it is so easy to click, students may be inclined to look up more words when an electronic dictionary is provided (Roby, 1991, cited in Chun, 2001) than when a conventional hard-copy dictionary is available. However, we do not agree that online dictionaries induce shallow processing. We assume that the level of word processing is not intrinsic to the type of dictionary. This is best illustrated with the chance a word was remembered if looked up because retention of the plus-relevant words was always higher than that of the minus-relevant target words. Since the same text, the same computer programme, and the same students were concerned, we can conclude that it is not the type of dictionary, paper or electronic, but the type of task that determines the level of word processing. As a result, the lower retention scores of the minus-relevant target words should not be attributed to the electronic dictionary but to the fact that these words were not processed as elaborately as the plus-relevant target words. However, our interpretation remains speculative because this study was not designed to compare paper and electronic dictionaries.

Pedagogical Implications

What are the pedagogical implications of this study? When reading a text, students are guided by a priority for meaning, even if they are forewarned of a vocabulary test. Our findings with respect to Word Relevance are very encouraging. Students recalled, dependent on the type of vocabulary test, two to three times more plus-relevant than minus-relevant words. Thus, we were successful in designing a task that directed students' attention to vocabulary in a text, which otherwise might have remained unnoticed. The reading comprehension task made the plus-relevant target words salient for learning. The fact that we made dictionary consultation easy contributed to the high retention scores of the plus-relevant words as well. In other words, it is possible to counterbalance students' natural tendency to mainly focus on text content by unobtrusively directing students' attention to words relevant in terms of the task. That the learners experienced this word-directing technique as unobtrusive was confirmed by the retrospective questions and think-aloud protocols which clearly show that the participants in this study were unaware of this pedagogic intervention.

CONCLUSION

This study aimed at finding empirical evidence for the effect of Test Announcement and Word Relevance on L2 learners' look-up behaviour and word retention. By having students carry out a reading comprehension task, we tried to direct their attention to unknown words in an unobtrusive way.

The results show that students' look-up behaviour is affected by Test Announcement as well as Word Relevance. In other words, it does make a difference when students know that a vocabulary test will follow. In addition, Task-induced Word Relevance also influences the frequency of lookups. The findings of this study also indicate that students forewarned of a vocabulary test look up more minus-relevant target words in comparison to students who are not forewarned.

Though Vocabulary Test Announcement has an effect on L2 learners' noticing of a target word, it does not affect their word retention. It does not change the way students approach reading a text because they are mainly driven by a priority for understanding the meaning of a text. Contrary to our expectations,

Vocabulary Test Announcement does not induce deeper levels of word processing, which is required if students want to learn L2 words. This study shows that vocabulary acquisition can be enhanced by providing L2 learners with word-directing reading comprehension questions. Such questions are a very efficient means to draw students' attention to unknown words. Our findings indicate that words deemed relevant to accomplish a task are better remembered. Moreover, the effect of Word Relevance does not diminish over time.

To conclude this paper, we would like to formulate two suggestions for future research. A first suggestion regards the online dictionary. We only provided a context-bound L1 translation and L2 explanation. Future research might explore how Test Announcement and Word Relevance would affect word retention when all meanings of (target) words are provided. Second, this study has demonstrated the effect of Word Relevance on retention of individual lexical items. However, because of the growing interest in multiword units, research is needed into the effect of this enhancement technique on retention of collocations or formulaic sequences as well.

NOTES

1. We will use the label *second language* (L2) for both second and foreign language.

2. The following words were annotated: all the words not belonging to the 2000 most frequent words in German, the 16 target words, and some lexical items of the 2000-word level that appeared to be problematic in the pilot study (Peters, 2006a, 2006b).

3. The questions were of the so-called "textually explicit" type, meaning that the question-answering process "involves search-and-match strategies rather than actual comprehension" (Alderson, 2001, p.107). Consequently, without access to the text, these questions become considerably more difficult to solve because the answers must be retrieved from memory.

4. This student gave the correct translations of two plus-relevant words. It was hardly possible to give the correct translation of *Pisaune* (authority) and *Fahnung* (sadness). No other student of the 84 participants managed to do this. This student might have copied off her neighbour. She was contrary to other students also able to translate ten words which she did not look up.

5. A recall test item requires the test-taker to supply the form or the meaning of the target word, whereas a recognition item involves the acceptance or rejection of the stimulus information. In the former test format, students have to provide either the L2 form or the meaning of the target word. In the latter test, students have to tick off the correct L2 form or meaning from among one or more options. They only need to recognize the correct option, which requires access (making a match) to, rather than retrieval from the mental lexicon (Laufer, Elder, Hill, & Congdon, 2004, p. 206; Nation, 2001, p. 359).

6. *IsolTest1*: Effect of Test Announcement F (1, 82) = 0.01; p = .92; η_p^2 = .00; Effect of Word Relevance F (1, 82) = 8.92; p <.01; η_p^2 = .10; Effect of interaction between two variables F (1, 82) = 0.13; p = .72; η_p^2 = .00

IsolTest2: Effect of Test Announcement F (1, 75) = 0.24; p = .63; η_p^2 = .00; Effect of Word Relevance F (1, 75) = 33.71; p <.0001; η_p^2 = .25; Effect of interaction between two variables F (1, 75) = 0.52; p = .48; η_p^2 = .07

IsolTest3: Effect of Test Announcement F (1, 53) = 14.14; p <.0001; η_p^2 = .22; Effect of Word Relevance F (1, 53) = 38.50; p <.0001; η_p^2 = .43; Effect of interaction between two variables F (1, 53) = 0.31; p = .58; η_p^2 = .01

7. *ContTest1*: Effect of Test Announcement F (1, 82) = 9.24; p = .33; η_p^2 = .01; Effect of Word Relevance F (1, 82) = 40.87; p < .0001; η_p^2 = .34); Effect of interaction between two variables F (1, 82) = 0.69; p = .41; η_p^2 = .01

ContTest2: Effect of Test Announcement F (1, 75) = 0.91; p = .35; η_p^2 = .01; Effect of Word Relevance F (1, 75) = 47.52; p < .0001; η_p^2 = .39; Effect of interaction between two variables F (1, 75) = 0.33; p = .57; η_p^2 = .01

ContTest3: Effect of Test Announcement F (1, 54) = 0.14; p = .71; η_p^2 = .00; Effect of Word Relevance F (1, 54) = 16.42; p < .0001; η_p^2 = .24; Effect of interaction between two variables F (1, 54) = 0.18; p = .67; η_p^2 = .00

8. The analysis of the ContTest and MatchTest was based on one observation more in comparison to the IsolTest.

9. *MatchTest1*: Effect of Test Announcement F (1, 82) = 1.63; p = .21; η_p^2 = .02; Effect of Word Relevance F (1, 82) = 0.33; p = .57; η_p^2 = .01; Effect of interaction between two variables F (1, 82) = 0.893; p = .35; η_p^2 = .01

MatchTest2: Effect of Test Announcement F (1, 75) = 1.93; p = .17; η_p^2 = .03; Effect of Word Relevance F (1, 75) = 14.07; p < .0001; η_p^2 = .16; Effect of interaction between two variables F (1, 75) = 1.58; p = .21; η_p^2 = .02

MatchTest3: Effect of Test Announcement F (1, 54) = 0.81; p = .37; η_p^2 = .02; Effect of Word Relevance F (1, 54) = 7.65; p = .008; η_p^2 = .13; Effect of interaction between two variables F (1, 54) = 0.60; p = .44; η_p^2 = .01

ACKNOWLEDGMENTS

I would like to thank the following persons for having me carry out the experiment in their classes: Dominique Diricken and Madeline Lutjeharms (Vrije Universiteit Brussels), Jan Van Nieuwenbergh (Erasmushogeschool, Brussels), Hans-Werner am Zehnhoff and Birgitta Meex (Lessius Hogeschool, Antwerp), and Marc Van de Velde and Michael Hinderdael (Hogeschool Gent). I am also greatly indebted to Arnoud Wils (K.U.Leuven) for developing the computer programme and to An Carbonez (K.U.Leuven) for her statistical advice. Since this study was carried out for my PhD, I would like to express my gratitude to my PhD-supervisors Lies Sercu (K.U.Leuven) and Jan Hulstijn (University of Amsterdam) for their invaluable comments on the design of this experiment. Finally, I would like to thank the anonymous reviewers for their very useful comments and suggestions on an earlier draft of this paper.

ABOUT THE AUTHOR

Elke Peters (http://wwwling.arts.kuleuven.be/flae/members/elke_peters.htm) obtained her PhD in 2006 at the K.U.Leuven (http://www.kuleuven.be). Her research focuses on second language acquisition and on vocabulary acquisition in particular.

E-mail: elke.peters@arts.kuleuven.Be

REFERENCES

Alderson, J.C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: a comparative study. *Language Learning & Technology*, *5*, 202-232. Retrieved January 31, 2001, from http://llt.msu.edu/vol5num1/alseghayer/default.html

Baddeley, A. (1997). *Human memory: Theory and practice* (Revised edition). Hove: Psychology Press.

Chun, D. (2001). L2 reading on the web: Strategies for accessing information in hypermedia. *Computer Assisted Language Learning, 14,* 367-403.

Chun, D., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, *80*, 183-198.

De Ridder, I. (2002). Visible or invisible links: Does the highlighting of hyperlinks affect incidental vocabulary learning, text comprehension, and the reading process? *Language Learning & Technology, 6*, 123-146. Retrieved January 21, 2003, from http://llt.msu.edu/vol6num1/DERIDDER/default.html

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21,* 303-317.

Hill, M., & Laufer, B. (2003). Type of task, time-on-task, and electronic dictionaries in incidental vocabulary acquisition. *IRAL*, *41*, 87-106.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*, 207–223.

Horst, M., Cobb, T., & Nicolae, I. (2005). Expanding academic vocabulary with an interactive on-line database. *Language Learning & Technology*, *9*(2), 90-110. Retrieved June 5, 2005, from http://llt.msu.edu/vol9num2/horst/default.html.

Hulstijn, J. H. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *The Modern Language Journal, 77*, 139-147.

Hulstijn, J. H. (2000). The use of computer technology in experimental studies of second language acquisition: A survey of some techniques and some ongoing studies. *Language Learning & Technology, 3*(2), 32-43. Retrieved October 29, 2002, from http://llt.msu.edu/vol3num2/hulstijn/index.html

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal, and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. Doughty & M.H. Long (Eds.), *Handbook of second language acquisition* (pp. 349-381). Malden, MA: Blackwell.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal, 80*, 327-339.

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

Institut für Deutsche Sprache - Mannheim (http://www.ids-mannheim.de)

Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal, 78*, 285-299.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing, 21*, 202-226.

Laufer, B., & Girsai, N. (2005, September). Vocabulary acquisition through text-based translation tasks. Paper presented at the annual EUROSLA conference, Dubrovnik, Croatia.

Laufer B., & Hill M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning & Technology*, *3*(2), 58-76. Retrieved October 29, 2002, from http://llt.msu.edu/vol3num2/laufer-hill/index.html

Laufer, B., & Hulstijn, J.H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*, 1-26.

Laufer, B., & Levitzky-Aviad, T. (2003). Look up behaviour and word retention as a function of task type and word relevance. *AsiaLEX Proceedings*, Tokyo, Japan.

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

Peters, E. (2006a). L2 vocabulary acquisition and reading comprehension: The influence of task complexity. In M. P. Garcia-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 178-198). Clevedon: Multilingual Matters.

Peters, E. (2006b). *Learning second language vocabulary through reading. Three empirical studies into the effect of enhancement techniques*. Katholieke Universiteit Leuven, Department of Linguistics, Unpublished doctoral dissertation.

QuestionMark Perception [Computer Software]. http://www.questionmark.com/index.htm

Rieder, A. (1998). Vokabellernen unter Multimediabedingungen: Erfahrungen aus dem TELOS-Projekt. In M. Auer, & U. Ressler (Eds.), *Interactive computer aided learning: Concepts and applications* (pp. 107-112). Klagenfurt: Kärntner Druckerei.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Sercu, L., De Wachter, L., Peters, E., Kuiken, F., & Vedder, I. (2006). The effect of task complexity and task conditions on foreign language development and performance. Three empirical studies. *ITL, International Journal of Applied Linguistics, 152*, 55-84.

Staud, T. (2003 October). Ossis sind Türken. 13 Jahre Einheit: In Gesamt-Westdeutschland sind die Ostdeutschen Einwanderer. *Die Zeit*, *41*. Retrieved October 2, 2003, from http://www.zeit.de/2003/41/Einwanderer

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, *19*, 287-307.