

Sentence Reading and Writing for Second Language Vocabulary Acquisition

^{1,*}FRANÇOIS PICHETTE, ²LINDA DE SERRES and
³MARC LAFONTAINE

¹Université du Québec, ²Université du Québec à Trois-Rivières and ³Université Laval, Canada

*E-mail: pichette.francois@teluq.ca

This study compares the relative effectiveness of reading and writing sentences for the incidental acquisition of new vocabulary in a second language. It also examines if recall varies according to the concreteness of target words. Participants were 203 French-speaking intermediate and advanced English as second language (ESL) learners, tested for incidental acquisition of 16 rare concrete, or abstract L2 words. Immediate and delayed cued recall was used to assess acquisition. Results from immediate recall show superior recall for writing tasks over reading tasks, and for concrete words over abstract words. However, delayed recall scores suggest that this superiority disappears over time.

Given its implications for teaching, a subject that has sparked interest is the comparative effectiveness of reading¹ and writing in the acquisition and retention of words in a second language (L2). Opinions differ regarding which individual activity is more likely to promote the retention of a new word by the learner: Is it reading a word in context or writing that word in a sentence? Unlike reading, which offers external input, writing is a language generating task, thus it does not allow for encountering new words. Therefore, the question of the relative efficiency of reading versus writing must be addressed regarding new words recently encountered by—or presented to—the L2 learner.

STATE OF THE QUESTION

Studies comparing the effectiveness of writing and reading tasks for the acquisition of vocabulary in L2 are scarce. In several cases, the writing of sentences was compared with the reading/visualizing of isolated words (e.g. Coomber *et al.* 1986; Barcroft 1998, 2000), or the reading/visualizing of words was compared with the writing of isolated words (e.g. Thomas and Dieter 1987; Barcroft 2006). However, reading or writing isolated words out of context is not a normal learning task. Furthermore, many of these studies opposed tasks carried out with sentences whose context offered semantic or syntactic information with tasks based on isolated words with no contextual clues.

Considering these limitations, other more recent studies have compared the reading and writing of plain text. Studies conducted by Hulstijn and Laufer (2001), Laufer (2003), Keating (2008), and Kim (2008) all compared the recall of words or pseudo-words in English as second language (ESL) after the completion of tasks relating to the reading and writing of sentences, passages or short texts. Contrary to the aforementioned studies relating to isolated words (Barcroft 2006), these studies reported a significantly higher recall for writing tasks. However, Keating (2008) notes that the superiority of writing in his study did not hold when test scores were adjusted to reflect time on task. In the first experiment of his two-part study, Webb (2005) had one group of participants write sentences containing pseudo-words while another group read the sentences with equivalents given in L1. Participants were given 12 min to learn the words. Although reading proved to be superior on 10 different recall measures, Webb points out that the readers had to wait for the slower writers to complete the task: the readers thus had several minutes to resort to memorization strategies after their task was complete, giving them a likely advantage over the writers. For Webb's second experiment, each participant completed both the reading and writing tasks, each with 20 pseudo-words. This time, however, they were given only the time necessary to complete each task. Lastly, participants were unaware that there would be a recall test, in order to avoid the temptation to resort to individual memorization strategies. The results ran contrary to the first experiment: writing proved to be superior on all recall measures.

The scarcity of these studies, compounded by their methodological discrepancies, makes it impossible to draw solid conclusions, so that current knowledge does not allow for definitive answers regarding the relative effectiveness of reading and writing for L2 vocabulary acquisition. Some studies and models suggest that reading is the more effective of the two activities, while others seem to conclude the opposite. Furthermore, researchers do not agree as to whether the higher level of complexity inherent to writing² tasks enhances acquisition. As summarized by Kuiken and Vedder (2008),

Skehan and Foster's Limited Capacity Model predicts better performance on the non-complex task as learners do not have to direct almost their entire attentional capacity towards the content of the task, and, thus, part of their attention can be devoted to the linguistic form. Robinson's Cognition Hypothesis, on the contrary, expects learners to do better on the complex task as learners are able to share their attention between content and form. (p. 51)

HYPOTHESES

In light of the above considerations, the goal of the present study is to investigate learning that is as incidental as possible,³ since the more intentional the

learning is, the more the learner is likely to resort to diverse memorization strategies. In such cases, one ends up comparing reading+strategies to writing+strategies. Our hypotheses are that writing tasks will better promote vocabulary acquisition and that concrete words will be better acquired.

Hypothesis #1: superiority of writing tasks

According to Swain's (1985) Output Hypothesis, only language production tasks truly compel the learner to undertake full grammatical processing, which favors a more effective L2 development (Gass 1988; Joe 1998). It is, in fact, generally accepted that word memory depends largely on both the amount and quality of attention that has been paid to various aspects of that word (cf. Hulstijn 2001). These considerations are reminiscent of Craik and Lockhart's (1972) *Levels of Processing Theory*. These researchers had originally hypothesized that writing would be more effective, since it would necessitate 'deeper' processing of new words.⁴

In addition, researchers have paid little attention to syntax in their models of vocabulary acquisition, which usually focus on word meanings and forms. Syntactic representations are usually seen as lexical projections, thus stemming from the acquisition of a word rather than leading to it (e.g. Vanniarajan 1997). For example, an existing vocabulary acquisition model is the TOPRA model (Barcroft 2000), which is based on the form and the meaning of new words, with syntax being absent as an influential factor. Barcroft (2004) emphasizes that:

Some researchers have posited that the type of semantic elaboration involved in sentence writing should facilitate learning new words (e.g. target pseudowords; Coomber *et al.* 1986). Other researchers have argued that the type of output involved in sentence writing can facilitate lexical learning (Laufer 1997). (p. 304)

Barcroft argues against a superiority of writing tasks for learning new words, claiming that language production relies on processes other than those involved in the acquisition of new forms, thus tapping the cognitive resources which are necessary (but no longer available) for acquisition. However, it is difficult to determine if the amount of semantic processing which takes place when encountering a new word is really superior when writing it in a complete sentence rather than when reading it in context, even when accompanied by an image or an L1 equivalent as semantic support. Also, we must not ignore an additional type of processing, which further distinguishes the writing of a sentence from the writing of isolated words, or from any type of reading: the *syntactic* elaboration,⁵ which occurs (in addition to the semantic elaboration) when combining words and appropriately placing them in a cohesive sentence structure.

Lastly, one further argument may be made in favor of the efficiency of writing over reading: the effect of task overlap. While writing necessarily involves a certain amount of reading, the contrary is not true. As we write, we visualize and read what we have written; would it not follow, then, that some of the advantages linked to visual contact that we attribute to reading are also likely to play a role in the case of writing?

In light of these considerations, our study is based upon the hypothesis that the greater cognitive demands of writing should lead to better acquisition of new words through writing than through reading text.⁶

Hypothesis #2: influence of concreteness

On one hand, in many models of vocabulary acquisition (and the studies that inspired them), the factors that seem best to predict recall are centered on the strategies, attitudes or motivation of the reader (e.g. Tseng and Schmitt's model 2008)—as well as the attention given to these factors by the reader and the context of the target word. Few models, if any, consider the factor of word concreteness, even though the superiority of concrete words for recall over abstract words has been demonstrated in numerous studies over the years (Paivio 1983; De Groot and Keijzer 2000; Hamilton and Rajaram 2001; Peters and Daum 2008). For example, Vanniarajan's (1997) model only considers syntactic category, morphology and case among word-related factors, and 'the *Involvement Load Hypothesis* assumes that word complexity factors are assumed to be held constant' (Hulstijn and Laufer 2001: 554).

Consequently, our second hypothesis is that the nature of a word plays a role in its retention. We thus suggest that concrete words have a higher potential for retention due to the higher number of connections they allow⁷ and offer additional connections through mental imagery (Paivio 1983).⁸

Drawing from Craik and Lockhart's (1972) notion of depth of processing, Hulstijn and Laufer's *Involvement Load Hypothesis* mentioned earlier proposes that the effectiveness of a language learning task rests on the learner's level of cognitive involvement, which is operationalized by the combination of three components: need, search and evaluation (Hulstijn and Laufer 2001; Laufer and Hulstijn 2001). While the *Involvement Load Hypothesis* has received considerable empirical support (for recent overviews of such support, see Keating 2008 and Kim 2008), evidence for the impact of word concreteness renders questionable the fact that this hypothesis holds word factors constant. In evaluating the language learning potential of a task, it seems justified to consider word complexity factors in addition to task-induced involvement. As a first step toward the goal of complementing the *Involvement Load Hypothesis*, this study investigated the relevance of considering word concreteness in the equation.

RESEARCH QUESTIONS

The present study attempts to answer the following research questions:

Question #1

For intermediate and advanced L2 students, does sentence writing lead to higher vocabulary gains relative to sentence reading?

Question #2

For intermediate and advanced L2 students, does recall vary according to the concreteness of target words?

Question #3

Does the impact of task and concreteness change over time?

Based on our two hypotheses, our predictions for recall are that writing will lead to higher recall than reading, and recall will be superior for more concrete words.

METHODS

Participants

The participants are 203 French-speaking ESL students, enrolled in a University in Quebec, Canada. The participants' mean age was 24.2 years, with a range of 18–53 years.

In order to ensure that the participants had a sufficient mastery in the L2 to complete the writing tasks, and to preclude the possibility that insufficient L2 competence would saturate working memory and prevent vocabulary acquisition, only intermediate and advanced students were tested.

Items

Since the participants had a relatively high level of competence in L2, we chose to use eight concrete words and eight abstract words which we considered to be very rare. Each word contained three syllables, and had no French cognates. For reasons of ecological validity and authenticity, real words were preferred to pseudo-words. The use of rare words made it almost impossible for the participants to have encountered them previously. As with pseudo-words, it also eliminates the need to quantify or qualify previous knowledge of each item by each participant, which is haphazard (Read 2000).⁹ The words were assumed to be as close as possible to the two extremities of Paivio *et al.*'s (1968) Likert scale for concreteness and imagery. This was verified in a related study ($N = 20$) based on a 7-point scale, in which our abstract words yielded a mean of 3.0 for concreteness and 3.3 for mental imagery, whereas means for the concrete

words were 5.8 and 4.9, respectively (see Supplementary Appendix SA for all 16 items).

Tasks

For the writing task, participants were to write three sentences per item, with each sentence containing the target L2 word. The target word, accompanied by its definition in the L1, preceded each sequence of three sentences.¹⁰

The reading task involved three sentences containing the target word in three different syntactic functions: as a subject, as a direct object, and as an indirect object. When possible, the target words were used twice in the singular and once in the plural form (see Supplementary Appendix SB for task samples). The use of relatively long sentences was seen as a compromise between isolated words and texts, which would have involved considerably more language material than the writing tasks they were compared with.

The recall task chosen was cued recall, which requires the participants to provide the L2 word via a clue offered by the experimenter. The measured knowledge is thus of a productive, not receptive, nature. Cued recall is recognized as sensitive to word forms, since the person tested does not have to recognize the L2 form, but retrieve it from memory and produce it correctly. Since the experiment included abstract words, the clues were L1 French definitions, since the use of illustrations would be difficult, if not impossible.

Time allotted for completing the tasks

Almost all experimenters in previous studies of this nature have imposed fixed amounts of time on the participants. As discussed by Webb (2005), however, this methodological approach throws into doubt the nature of the processes that actually take place during the experiment, for several reasons.

First, if more time than necessary is allocated for a task, this leaves time for processes and learning strategies other than those anticipated by the researcher. Reading speed varies from one person to another, and once a sentence or clause has been read, the reader's eyes may remain fixed on the items; this has been called the *wrap-up effect*, which results in additional comprehension processes in the form of elaborative inferences (Wiley and Rayner 2000). Upon completion of these stages, the goal of comprehension is reached. While reading should be the only task involved, there are reasons to believe that if the reader is exposed to the same sentence longer than necessary, additional information processing will occur: re-reading, mental rehearsals, mnemonic techniques (in the case of intentional acquisition), etc. It is most likely that these strategies, both in nature and number, vary from one person to another. As signaled by Webb (2005), in a normal reading situation, nobody spends more time than necessary for completing a reading task.

The same phenomenon occurs in writing. The creation of a mental representation followed by marking it on paper might require different amounts

of time from one person to another. In this case, too, allocating a period of time longer than necessary will lead to cognitive processes in addition to the writing task.

In the same line of thought, too little time allotted could also have a negative effect on recall. In a normal reading or writing situation, nobody is subjected to a pre-determined, pre-announced number of seconds for completing the task. In studies with a fixed time, some participants struggle to complete the task on time (e.g. Webb 2005, Exp #1). In other cases, it is not unreasonable to assume that even if the time allocated was just enough (e.g. 48 s per sentence in Barcroft 2000), the mere awareness of this limitation could create stress which would be detrimental to performance. Therefore, controlling for time on task was deemed inadequate in this study given the different nature of the tasks involved, although it is advocated for tasks of a more similar nature (cf. Folse 2006 for different types of writing tasks). In line with the aforementioned ecological validity argument, these more 'natural' conditions akin to most classroom tasks may give our results more applicability for teaching.

Thus, as in Webb's (2005) second experiment, each individual participant was allowed the time necessary to complete the task; no time limit was announced. The writing task lasted for about twice as long as the reading task. This greater time, according to Webb (2005) and Hulstijn and Laufer (2001), is seen not as an advantage of the writing task, but rather as an intrinsic characteristic of it.

Procedure

The experiment was presented as a study aimed at measuring skills for comprehending and using rare words in English. It was decided not to disclose the real objective of the exercise so as to avoid the use of other idiosyncratic strategies for the intentional memorization of content. The participants were not made aware of the immediately following recall test; this was to promote incidental learning by reducing the risk that they focus on the target words for deliberate retention (Peters 2007).

The participants first signed a consent form and completed an identification questionnaire to gather general information regarding their literacy. Participants also received guidelines on the nature of the experiment with an example for each task.

The test consisted of one target word per sheet, displayed on the front side. This page was to be turned over as soon as it was completed. For the writing tasks, participants were instructed verbally to write sentences of more than 10 words, and to use a different type of sentence each time. This last requirement was among the instructions given orally at the time of testing, in order to limit the amount of written instructions on the test package. In most cases, the participants wrote the word as a subject, a direct object and an indirect object, as was shown in class as an example prior to testing (cf. the writing sample in

Supplementary Appendix SB). For the reading task, they were instructed to read each sentence once and to focus on understanding it.

This study consisted of four conditions with four items per condition: writing concrete words, writing abstract words, reading concrete words, reading abstract words. A within-subjects design was adopted. Each participant received all 16 target words, with eight randomly selected words to write and the other eight to read. Each word received an equal amount of each of the two treatments. Of the eight items, the participants saw in each task (writing and reading), half of those were abstract and half were concrete. Processing each word in both formats—reading and writing—by each participant was rejected, since for the second task the word would not be new anymore. The order of presentation of the items was randomized, and the order of the tasks was balanced: tasks alternated between reading and writing for the duration of the exercise.

Due to the short testing time, the tests were administered during normal class time. A financial compensation was offered in the form of a lottery.

The surprise recall task was given to each participant individually almost immediately after he or she completed the tasks: this consisted of a sheet with the L1 definitions for the 16 target words, which they had 2 min to write down. At the bottom of the same sheet, participants were asked to indicate which words, if any, they had already known prior to the experiment. The same test was administered a week later in order to measure longer term recall.

Scoring and analyses

The scores were calculated by an independent corrector and the participants were kept anonymous. One point was awarded for each correct syllable, with a maximum possible score of three per word. Syllable scoring was chosen due to its high correlations with scores resulting from whole word scoring (e.g. $r = .96$ in Farhadi and Malekpour 1997). It also allows correctors not to reject forms such as *opsimat* or *opsimoth* for *opsimath*, all the while providing them with an objective method for allowing partial points. Significant effects can be observed based on syllable scoring that may not be visible for whole word scoring, due to a lack of precision of the latter as a measure of word form knowledge, as hypothesized by Barcroft (2004). A correct form associated with a wrong word was counted solely on the basis of formal similarities with the correct L2 word. Any word that a participant had already seen was excluded from the analysis for that participant, as well as words for which the task was not completed according to the directions given. The scores were put in a MS Access data base. The analysis was conducted using SAS 9.2, with mixed-model repeated measures.

Among the few studies cited that compared reading and writing of sentences or text, with sufficient time allowed for performing the tasks, only one researcher (Webb 2005, 2nd experiment) used a within-subjects design, that is where participants were exposed to all task types, which ‘decreased the

possibility of individual differences influencing the results' (p. 45). However, in this case as in all the studies listed that met the above-mentioned criteria, the variable of word concreteness was not considered; only word categories are sometimes mentioned, but with no comparative analyses based on those categories. Therefore, given the small number of items generally tested for recall, the relative effectiveness of the two types of tasks may depend not only on the nature of the task, but also on the concreteness of the items whose acquisition is assessed. One of the major contributions of the present study is to separate out the variables related to the task performed and those related to the items learned, among which concreteness has been shown to influence word retention, as was emphasized earlier.

RESULTS

The testing sessions proceeded as planned, with testing times ranging from 18 to 34 min, with an average of 24 min. The writing tasks seemed to take about three times longer than the reading task, as opposed to twice as long in Webb (2005). Of the 203 completed test packages, only one yielded a missing task, which was the writing task for *treppverter*. In total, each word was processed between 100 and 105 times for each of the two tasks. Of a total of 3248 items (words \times participants), 66, or 2.0 per cent, were marked as having been encountered before. About half the cases were for the word *buoyancy* ($N = 30$), followed by *acronyx* ($N = 6$) (see Supplementary Appendix SB). The remaining cases are shared about equally among the rest of the words, all with zero to four claims of previous encounter.

Given the very rare nature of the words, this impression is likely to rest solely on vague *déjà-vu* based on resemblance with other words, and similar factors, instead of actual prior encounters with said words. As support for this hypothesis, Participant #45 claimed to have previously encountered half the words (!), and at Recall #1 marked *fustilug* as having been encountered, but did not do so at Recall #2. Likewise, both P#63 and P#88 identified three words as having been encountered before the test, which is very unlikely to be the case.⁹ Scores for the rejected items, without including the word *Buoyancy*, were 1.17 out of 3 (SD 1.40) for Recall #1 and 1.04 out of 3 (SD 1.43) for Recall #2. These figures are higher than the means obtained for the data we analyzed, as will be seen in this section, while not as high as one would expect if all these words had been known to the participants. On the contrary, *Buoyancy* seems to be a less uncommon word, and had most likely been encountered, since recall for this item is 2.80 (SD 0.55) for Recall #1 and 2.67 (SD 0.87) for Recall #2.

The scores presented in this section underwent square root transformation in order to reach a distribution closer to normal, using the BoxCox family of transformations. Variance postulates were verified to ensure normality and homogeneity on model residues, that is transformed data. Since normality was not entirely satisfactory, the analysis was also conducted on ranks, that

Table 1: Scores out of 3, task versus recall

	Reading task		Writing task	
	Raw data, mean (SE)	Transformed data, mean (SE)	Raw data, mean (SE)	Transformed data, mean (SE)
Recall 1	0.4718 (0.0350)	0.4887 (0.0282)	0.7645 (0.0457)	0.6937 (0.0309)
Recall 2	0.2371 (0.0285)	0.2590 (0.0249)	0.3150 (0.0346)	0.3173 (0.0285)

Table 2: Tests of effect slices, task versus recall

Effect	Recall	<i>df</i>	<i>F</i>	<i>p</i>
Recall*task	1	202	47.28	<.0001
Recall*task	2	175	4.59	.0335

is non-parametric values. The analysis was also repeated on raw, not transformed data. In all three cases the same conclusions are reached, although *p*-values differ slightly.

Question #1

For intermediate and advanced L2 students, does sentence writing lead to higher vocabulary gains relative to sentence reading?

Table 1 shows the scores out of 3 for both tasks (reading and writing) and for both recall tests (immediate and delayed).

The *F*-tests show a significant difference in the average scores for both recall tests. More specifically, the average score is lower for words read compared with words written for both immediate and delayed recall tests.

Furthermore, additional *F*-tests have shown that the average scores for both recall tests are significantly lower for the second (delayed) recall, both for words read ($F=116,13$, $p < .0001$) and words written ($F=190, 90$, $p < .0001$).

Table 2 shows a recall/task interaction near the α -level for Recall 2. A closer look through differences of Least Square Means (LSMs) shows that significance is reached due to the difference between recall for concrete words in reading and writing ($t=-2.41$, $df=173$, $p=.0172$). There would be no difference between reading and writing had we tested abstract words only ($t=-0.71$, $df=171$, $p=.4760$).

Table 3: Scores out of 3, concreteness versus recall

	Abstract words		Concrete words	
	Raw data, mean (SE)	Transformed data, mean (SE)	Raw data, mean (SE)	Transformed data, mean (SE)
Recall 1	0.5619 (0.0279)	0.5697 (0.0365)	0.6205 (0.0304)	0.6666 (0.0427)
Recall 2	0.3010 (0.0257)	0.2810 (0.0298)	0.2753 (0.0265)	0.2711 (0.0328)

Table 4: Tests of effect slices, concreteness versus recall

Effect	Recall	df	F	p
Recall*task	1	201	4.37	.0379
Recall*task	2	156	1.08	.3002

Question #2

Does recall vary according to the concreteness of target words?

Table 3 shows the scores for both word types (abstract and concrete) and for both recall tests (immediate and delayed).

As shown in Table 4, *F*-tests show a significant difference between both word types (abstract and concrete) for immediate recall only ($p = .0379$). More precisely, the average score is lower for abstract words than for concrete words on immediate recall tests. There was no significant difference between the average scores of abstract or concrete words on the delayed recall tests ($p = .3002$).

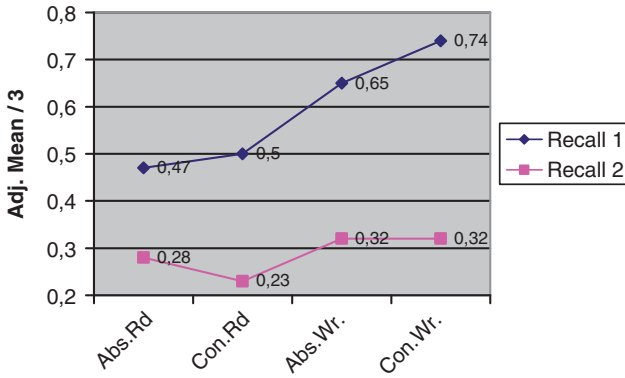
Also, *F*-tests show that the average scores for both types of recall were significantly lower for the second recall test, whether it be concrete words ($F = -173.27$, $p < .0001$) or abstract words ($F = 136.29$, $p < .0001$).

Still in Table 4, the only significant recall/task interaction, that for Recall 1, warrants further analyses. Differences of LSMs show that significance is reached due to the writing task ($t = -2.07$, $df = 201$, $p = .0394$), the reading task showing no difference between concrete and abstract words ($t = -0.80$, $df = 201$, $p = .4246$).

Question #3

Does the impact of task and concreteness change over time?

In order to address Question 3, the variables of task and concreteness were combined, and the two recall tasks were separated. As Figure 1 shows, a comparison between scores for each combination shows an interesting leveling



Note: Abs = abstract; Con = concrete; Rd = reading; Wr = writing.

Figure 1: Adjusted mean scores for all four combinations—task versus concreteness

trend occurring between Recall 1 and Recall 2. For Recall 1, differences in LSMs are significant for all combinations except between the reading of abstract and the reading of concrete words ($p = .4246$). For Recall 2, the significant differences stem only from the reading of concrete words, whose adjusted mean score is significantly lower than for the writing of concrete words ($p = .0172$) and of abstract words ($p = .0178$).

DISCUSSION

The data profile obtained matched our predictions and expectations based on Webb (2005) and on the Involvement Load Hypothesis, namely that, generally, writing a text may lead to significantly higher recall than reading if enough time is allocated for each task, writing being intrinsically longer than reading for the same amount of language.

Furthermore, observed recall was superior for more concrete words. This is yet further empirical support for this robust phenomenon which has been observed and studied for several decades.

Concerning the length of the observed recall, it is pertinent to compare the results of the second recall with that of the first. As for the apparent superiority of writing tasks, this difference seems to hold steady, at least for 1 week, since it persists in the second recall. However, our analyses showed that by the second recall, it had become borderline significant, having disappeared for abstract words. As for the superiority of concrete over abstract words, this difference quickly waned in the present study, as it did not appear in the delayed recall test. The high amount of empirical support for the concreteness superiority effect from previous research is probably related in part to the high number of

studies that tested recall a short period of time after the experiment. With a more longitudinal version of this study, it would be interesting to see if this leveling pattern would continue over a longer period, and if this apparent superiority for writing would continue to degrade up to the point of disappearing completely; and, if so, after what length of time.

If we assume that delayed recall is a better indicator of learning than immediate recall, the prediction that writing will lead to higher vocabulary retention than reading seems to be supported only for concrete words and not for abstract words when recall is assessed after 1 week. From a theoretical standpoint, this intriguing observation has important implications for the Involvement Load Hypothesis. In addition to suggesting that word factors should be taken into consideration when assessing the learning potential of a language task, it underscores the need for investigations into why learners show little or no benefits of writing activities over reading activities when it comes to language items of lower concreteness.

Pedagogical implications

Writing short passages for proficient L2 students can be an effective activity for retaining words that have recently been encountered. However, it has to be kept in mind that the effectiveness of teaching methods varies according to learner types, and teaching outcomes depend mostly on the motivating aspects that stem from using a variety of activities. Nevertheless, our results suggest that language teachers may resort to writing tasks that incorporate newly taught words in order to enhance students' retention.

Limitations

Among the three aspects of a word knowledge according to Nation (2001) (form, meaning, and use), only the first two aspects were involved in the use of our reading and writing tasks. A separate measure for the grammatical functions of target words would perhaps shed more light on the relative impact that the two types of tasks would have on word retention. Examining a variety of writing tasks would be another interesting avenue to explore (Joe 1998; Waring and Tataki 2003; Folse 2006). It is also hoped that future studies will find a way to measure the amount of syntactic and semantic processing in reading and writing, to see if that variable plays a role in determining their relative efficiency for acquiring language.

Moreover, this study aimed at measuring recall which was to be as incidental as possible. It remains to be seen whether the incidental nature of the experiment and our data could have been influenced by two factors. First, target words were displayed on the top of the test pages as well as during the task, thus creating some overemphasis on those items. Second, many students had probably guessed that some sort of recall test would be given as a follow-up task, given the fact that they are frequently solicited for participating in studies

during their degree program. And there remains the ever-present question of the extent to which recall measures actually reflect acquisition.

CONCLUSION

In light of these considerations, and taking into account the above-mentioned studies, many researchers argue that incidental vocabulary acquisition through reading is limited (e.g. Pichette 2005; Pigada and Schmitt 2006; Esquiliche Mesa *et al.* 2007; Pulido 2007; Brown *et al.* 2008). We have reason to believe that a purely incidental learning situation would have led to even lower recall scores. This study nonetheless suggests that some limited incidental acquisition can take place during 'normal' reading and writing activities, and that both exposure to and production of language can lead to some (albeit very limited) incidental retention. We hope that our study—and others like it—may benefit teachers and pedagogues seeking to maximize their class time and their students' language learning potential.

SUPPLEMENTARY DATA

Supplementary Data are available at *Applied Linguistics* online.

ACKNOWLEDGEMENT

The authors would like to express our gratitude to the four anonymous reviewers for their insightful suggestions that led to improvements in this article.

NOTES

- 1 In the present study, writing is seen as the production in the written form of plain text, coming from the writer and significant to him/her. It is not a question of isolated words, copied words, or words unfamiliar to the writer. Reading, on the other hand, is seen as a quest for meaning through the visual perusal of plain text. This form of reading is the most commonly found and the generally accepted definition of *reading* in research ('reading for meaning'; Swaffar *et al.* 1991). It is thus not a question of visualizing isolated words, of scanning text, which consists of rapid glancing at a page in search of particular graphical symbols, or skimming text, which consists of locating only certain elements of the text in order to draw a general impression.
- 2 We acknowledge the fact that complexity depends not only on whether reading or writing is involved but on what is required of the learner: some writing tasks can be simple, and some reading tasks complex. The assumption here is that cognitively speaking, writing tasks will be generally more demanding than reading tasks.
- 3 For a discussion on the difference between intentional and incidental learning, see Hulstijn (2003).
- 4 Syntactic processing has always been regarded as a complex task that is only accomplished by humans, and

not by animals that produce 'language' (Pinker 1994), and on the condition that humans start before a critical age (Curtiss 1977). As for semantic processing, it is less complex, demands less effort and is manifested in some animal species though comprehension or even word production.

- 5 'Semantic elaboration refers to increased evaluation of an item with regard to its meaning' (Barcroft 2002: 323). Syntactic elaboration refers to increased evaluation of an item with regard to its syntactic properties. In this article, *elaboration* and *processing* are considered synonyms.
- 6 According to the Generation Effect described among others by McNamara and Healy (1995), it should be noted that the benefits of language production derive from the production of the sentence and not from the output stage. This means that it is not the writing of a sentence on paper that promotes acquisition, but the semantic and syntactic processes that precede it.
- 7 This hypothesis is closely linked to the *Context Availability Theory*, which assumes that concrete words activate a broader contextual verbal support, which is reflected by a greater

activation of left parietal and frontal associative areas (Jessen *et al.* 2000).

- 8 Paivio's (1983) Additivity Hypothesis, with considerable empirical support, argues that the additional presentation of an illustration, in addition to the word form itself, yields recall superior to that of the form alone, even though in theory more cognitive resources are necessary for a double treatment than for a single one.
- 9 We verified with five native speakers of English who had completed their doctoral studies; none of them was familiar with more than three words.
- 10 An earlier version of the test consisted of writing sentences on blank lines on which two words were already provided: the target words as well as a verb that would force the participant to adopt the same three syntactic structures for the reading task. Tested with 122 participants, the absence of impact of the type of activity as well as of the item type (Pichette *et al.* 2008) coupled with participants' remarks underscored the importance for people to write the target words in the written task, with the risk of not seeing all three syntactic structures being produced. The present article reports on the use of this newer version of the writing tasks.

REFERENCES

- Barcroft, J.** 1998. *L2 Vocabulary Learning: Do Sentence Writing and Oral Repetition Help?* Presented at the Second Language Research Forum, October. Honolulu, HI.
- Barcroft, J.** 2000. *The Effect of Sentence Writing as Semantic Elaboration on the Allocation of Processing Resources and Second Language Lexical Acquisition*. Doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana.
- Barcroft, J.** 2002. 'Semantic and structural elaboration in L2 lexical acquisition,' *Language Learning* 52/2: 323–63.
- Barcroft, J.** 2004. 'Effects of sentence writing in second language lexical acquisition,' *Second Language Research* 20/4: 303–34.
- Barcroft, J.** 2006. 'Can writing a new word detract from learning it? More negative effect of forced output during vocabulary learning,' *Second Language Research* 22/4: 487–97.
- Brown, R., R. Waring, and S. Donkaewbua.** 2008. 'Incidental vocabulary acquisition from reading, reading while listening, and listening to stories,' *Reading in a Foreign Language* 20/2: 136–63.

- Coomber, J. E., D. A. Ramstad, and D. R. Sheets. 1986. 'Elaboration in vocabulary learning: A comparison of three rehearsal methods,' *Research in the Teaching of English* 20: 281-93.
- Craik, F. I. M. and R. S. Lockhart. 1972. 'Levels of processing: A framework for memory research,' *Journal of Verbal Learning and Verbal Behavior* 11: 671-84.
- Curtiss, S. 1977. *Genie: A Psycholinguistic Study of a Modern-Day "Wild Child"*. Academic Press.
- De Groot, A. M. B. and R. Keijzer. 2000. 'What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting,' *Language Learning* 50: 1-56.
- Esquiliche Mesa, R., A. Bruton, and T. Ridgway. 2007. 'The effect of task-based reading on FL vocabulary learning,' *ITL International Journal of Applied Linguistics* 153: 1-22.
- Farhadi, H. and M. Malekpour. 1997. 'Different methods of scoring and the psychometric characteristics of dictation tests,' *Journal of Social Sciences and Humanities* 12/1. Available at http://www.aau.academia.edu/dep/hf_publications/2%20Farhady.Malekpour.pdf. Accessed 9 September 2011.
- Folse, K. 2006. 'The effect of type of written exercise on L2 vocabulary retention,' *TESOL Quarterly* 40/2: 273-93.
- Gass, S. 1988. 'Integrating research areas: A framework for second language studies,' *Applied Linguistics* 9: 198-217.
- Hamilton, M. and S. Rajaram. 2001. 'The concreteness effect in implicit and explicit memory tests,' *Journal of Memory and Language* 44/1: 96-117.
- 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*. Cambridge University Press, pp. 258-86.
- Hulstijn, J. H. 2003. 'Incidental and intentional learning' in C. J. Doughty and M. H. Long (eds): *Handbook of Second Language Acquisition*. Blackwell, pp. 349-81.
- Hulstijn, J. H. and B. Laufer. 2001. 'Some empirical evidence for the involvement load hypothesis in vocabulary acquisition,' *Language Learning* 51: 539-58.
- Jessen, F., R. Heun, M. Erb, D. O. Granath, U. Klose, A. Pappasotiropoulos, and W. Grodd. 2000. 'The concreteness effect: Evidence for dual coding and context availability,' *Brain and Language* 74/1: 103-12.
- Joe, A. 1998. 'What effects do text-based tasks promoting generation have on incidental vocabulary acquisition?,' *Applied Linguistics* 19: 357-77.
- Keating, G. D. 2008. 'Task effectiveness and word learning in a second language: The involvement load hypothesis on trial,' *Language Teaching Research* 12/3: 365-86.
- Kim, Y. 2008. 'The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition,' *Language Learning* 58/2: 285-325.
- Kuiken, F. and I. Vedder. 2008. 'Cognitive task complexity and written output in Italian and French as a foreign language,' *Journal of Second Language Writing* 17/1: 48-60.
- Laufer, B. 1997. 'Incidental vocabulary acquisition: In praise of output.' Paper presented at the Second Language Research Forum, East Lansing, MI.
- Laufer, B. 2003. 'Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading?,' *Canadian Modern Language Review* 59/4: 565-585.
- Laufer, B. and J. H. Hulstijn. 2001. 'Incidental vocabulary acquisition in a second language: The construct of task-induced involvement,' *Applied Linguistics* 22: 1-26.
- McNamara, D. S. and A. E. Healy. 1995. 'A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems,' *Journal of Memory and Language* 43: 652-679.
- Nation, P. 2001. *Learning Vocabulary in Another Language*. Cambridge University Press.
- Paivio, A. 1983. 'The empirical case for dual coding' in J. C. Yuille (ed.): *Imagery, Memory and Cognition: Essays in Honor of Allan Paivio*, pp. 307-32.
- Paivio, A., J. C. Yuille, and S. A. Madigan. 1968. 'Concreteness, imagery, and meaningfulness values for 925 nouns,' *Journal of Experimental Psychology* 76: 1-25.
- Peters, E. 2007. 'Manipulating L2 learners' online dictionary use and its effect on L2 word retention?,' *Language Learning and Technology* 11/2: 36-58 Available at <http://llt.msu.edu/voll1num2/pdf/peters.Pdf>. Accessed 9 September 2011.

- Peters, J.** and **I. Daum.** 2008. 'Differential effects of normal aging on recollection of concrete and abstract words,' *Neuropsychology* 22/2: 255–61.
- Pichette, F.** 2005. 'Time spent on reading and reading comprehension in second language learning,' *Canadian Modern Language Review* 62/2: 243–62.
- Pichette, F., L. de Serres,** and **M. Lafontaine.** 2008. 'Word Concreteness and L2 Vocabulary Acquisition.' Paper presented at the AAAL Conference, Washington, DC.
- Pigada, M.** and **N. Schmitt.** 2006. 'Vocabulary acquisition from extensive reading: A case study,' *Reading in a Foreign Language* 18/1: 1–28.
- Pinker, S.** 1994. *The Language Instinct*. William Morrow & Co.
- Pulido, D.** 2007. 'The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity?,' *Language Learning* 57/1: 155–99.
- Read, J.** 2000. *Assessing Vocabulary*. Cambridge University Press.
- Swaffar, J. K., K. M. Arens,** and **H. Byrnes.** 1991. *Reading for Meaning. An Integrated Approach to Language Learning*. Prentice Hall.
- Swain, M.** 1985. 'Communicative competence: Some roles of comprehensible input and comprehensible output in its development' in S. Gass and C. Madden (eds): *Input in Second Language Acquisition*. Newbury House, pp. 235–53.
- Thomas, M. H.** and **J. N. Dieter.** 1987. 'The positive effects of writing practice on integration of foreign words in memory,' *Journal of Educational Psychology* 79: 249–53.
- Tseng, W.-T.** and **N. Schmitt.** 2008. 'Toward a model of motivated vocabulary learning: A structural equation modeling approach,' *Language Learning* 58/2: 357–400.
- Vanniarajan, S.** 1997. 'An interactive model of vocabulary acquisition,' *Applied Language Learning* 8/2: 183–216.
- Waring, R.** and **M. Takaki.** 2003. 'At what rate do learners learn and retain new vocabulary from reading a graded reader?,' *Reading in a Foreign Language* 15/2: 130–63.
- Webb, S.** 2005. 'Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge,' *Studies in Second Language Acquisition* 27/1: 33–52.
- Wiley, J.** and **K. Rayner.** 2000. 'Effects of titles on the processing of text and lexically ambiguous words: Evidence from eye movements,' *Memory and Cognition* 28: 1011–21.