Incidental vocabulary learning by advanced foreign-language students: the influence of marginal glosses, dictionary use, and reoccurrence of unknown words

Hulstijn, J.H.; Hollander, M.; Greidanus, T.

Published in:
The Modern Language Journal

DOI:
10.2307/329439

Incidental Vocabulary Learning by Advanced Foreign Language Students: The Influence of Marginal Glosses, Dictionary Use, and Reoccurrence of Unknown Words

Jan H. Hulstijn; Merel Hollander; Tine Greidanus

The Modern Language Journal, Vol. 80, No. 3 (Autumn, 1996), 327-339.

Stable URL:
http://links.jstor.org/sici?sici=0026-7902%28199623%2980%3A3%3C327%3AIVLBAF%3E2.0.CO%3B2-I

The Modern Language Journal is currently published by National Federation of Modern Language Teachers Associations.

Incidental Vocabulary Learning by Advanced Foreign Language Students: The Influence of Marginal Glosses, Dictionary Use, and Reoccurrence of Unknown Words

JAN H. HULSTIJN
Vrije Universiteit
Amsterdam, The Netherlands
Email: hulstijn@jet.let.vu.nl

MEREL HOLLANDER
Vrije Universiteit
Amsterdam, The Netherlands

TINE GREIDANUS
Vrije Universiteit
Amsterdam, The Netherlands

Dutch advanced students of French read a French short story in one of three text reading conditions: Marginal Glosses (provision of L1 translations of unknown words), Dictionary (opportunity to use a bilingual dictionary), or Control. After reading, students were tested for their recall of 16 words that had appeared either once or three times in the text. Support was found for the hypothesis that frequency of occurrence will foster incidental vocabulary learning more when advanced second language (L2) readers are given the meanings of unknown words through marginal glosses or when they look up meanings in a dictionary than when no external information concerning unknown words' meanings is available. In the former case, reappearance of a word will reinforce the form-meaning connection in the reader's mental lexicon. In the latter case, readers will often ignore unknown words or incorrectly infer their meanings, which will limit the frequency effect. This article ends with recommendations for teachers and researchers.

INTRODUCTION

IN THE LITERATURE ON FOREIGN OR SECond language (L2) instruction, it is a generally accepted principle that extensive L2 reading is good for vocabulary acquisition (Brown, 1994, chap. 16; Grabe, 1991; Nation, 1990; Swaffar, Arens, & Byrnes, 1991), just as first language (L1) reading, according to most researchers, leads to L1 vocabulary growth (Sternberg, 1987; Nagy, Anderson, & Herman, 1987).[1] It has been shown, at least for the Dutch language, that L2 learners, upon entering the university, have an average L2 receptive vocabulary knowledge of 11,000 words (Hazenberg & Hulstijn, 1996). It should be obvious that words in such large quantities cannot have been learned solely by means of intentional word-learning activities, that is, by committing word forms to memory along with their meanings. Surely, many words must have been "picked up" during listening and reading activities while the listener's or reader's goal was to comprehend the meaning of the language heard or read, rather than to learn new words. This "picking up" is usually referred to as incidental learning. Incidental learning can only be defined in negative terms as the accidental learning of information without the intention of remembering that information (Hulstijn, 1989; Schmidt, 1994).

On first sight then, it appears that vocabulary growth stems partly from reading and listening. On closer examination, however, it turns out that readers often fail to spontaneously learn the meanings of previously unknown words encountered in texts for one or several of the following reasons:

1. Sometimes, learners simply fail to notice the presence of unfamiliar words or believe that they know a word when, in fact, they do not.

The Modern Language Journal, 80, iii (1996)
0026-7902/96/327–339 $1.50/0

2. Sometimes, learners do notice the presence of unfamiliar words but decide to ignore them.

3. The contextual information may be so redundant that readers fail to connect the form of the unknown word to the meaning contained in the context. In other words, they may only pay attention to the meaning and ignore the unfamiliar word form. For learning to take place, however, attention must be focused not exclusively on the meaning of the target word, but also on the connection between the word's form and meaning. (For a discussion of this paradox, see Coady, 1993; Ellis, 1994; Nation & Coady, 1988; Watanabe, 1992.)

4. Often, the meanings of unknown words are not inferrable from the context. Readers also frequently make erroneous inferences and, hence, learn words incorrectly (Bensoussan & Laufer, 1984; Carnine, Kameenui, & Coyle, 1984; Dubin & Olshtain, 1993; Huckin & Haynes, 1993; Hulstijn, 1992; Laufer & Sim, 1985; Summers, 1988).

5. Even when readers have a dictionary available during reading, they do not look up all unknown words, especially when they are reading texts longer than a few hundred words (Hulstijn, 1993; Krantz, 1991). Perhaps L2 learners make more use of the dictionary when reading short texts. Knight (1994) found that L2 learners who had to read and retell in L1 a 250-word text made frequent use of the dictionary. In Knight's study, participants read the text on the screen of a computer monitor and accessed the dictionary via computer. However, when there is a new-item overload, regardless of whether a text is short or long, even good L2 learners soon abandon their intentions to look up new items (Jones, 1995).

6. Usually, a single encounter with a new word does not guarantee its acquisition. Nagy, Herman, and Anderson (1985), therefore, called the learning of words through reading a "process of small increments."

Several researchers have investigated which factors might promote incidental vocabulary learning. Their findings can be summarized as follows:

1. Deep elaboration on the meaning of an unknown word positively affects incidental learning (Hulstijn, 1992; Mondria & Wit-de Boer, 1991; Watanabe, 1992). Thus, inferred *meanings* are remembered slightly better than given meanings. However, this advantage turns into a disadvantage when readers infer a wrong meaning, which often happens to be the case (Hulstijn, 1992; Laufer & Sim, 1985; Schouten-van Parreren, 1989, 1992).

2. Readers tend to pay more attention to words whose meanings they consider to be relevant to an understanding of the text than to words not perceived as relevant (Hulstijn, 1993).

3. Readers with high verbal ability appear to "pick up" more words incidentally than readers with low verbal ability (Knight, 1994).

4. The use of a dictionary positively affects incidental vocabulary learning. In the Knight (1994) study, American students of Spanish who had read two Spanish texts while using a dictionary later remembered more word meanings than those who had had no dictionary at their disposal.

5. The provision of marginal vocabulary glosses not only enhances text comprehension (Davies, 1989), but also incidental vocabulary learning (Hulstijn, 1992; Jacobs, Dufon, & Fong, 1994; Watanabe, 1992). Learners with large vocabularies also profit more from marginal glosses than do learners with small vocabularies (Jacobs, Dufon, & Fong, 1994).

6. When the meanings of unknown words are given, either by marginal glosses or by dictionary use, L1 cues appear to have more effect than L2 cues (Krantz, 1991; Oskarsson, 1975; Scherfer, 1993). However, some learners prefer L2 glosses, if comprehensible, over L1 glosses (Jacobs, Dufon, & Fong, 1994).

7. Words appearing frequently in a text are more likely to be acquired than words occurring only once (Saragi, Nation, & Meister, 1978; Krantz, 1991). However, frequency alone does not determine the likelihood of acquisition. For instance, only 8 out of 20 subjects in the Clockwork Orange study by Saragi, Nation, and Meister (1978) knew the meaning of a word which had occurred 96 times in the novel!

In summary, research has shown that intermediate and advanced L2 learners "pick up" only few new, hitherto unfamiliar words by just reading for recreational purposes. Incidental vocabulary learning during reading does indeed take place but only incrementally and in small quantities. Thus, given the undeniable usefulness of L2 reading (e.g., for the improvement of automatic word recognition), the educationally relevant question remains how this generally low incidence of incidental vocabulary learning can be improved. Among the factors mentioned above that favorably affect incidental vocabulary learning are (a) the provision of marginal glosses, (b) the use of a dictionary, and (c) the reappearance or reoccurrence of new words in the text.

RESEARCH QUESTIONS

In the present study, we explored whether the reoccurrence of unknown words, in combination with the provision of information concerning their meanings, would increase the likelihood of incidental vocabulary learning. Thus, we investigated the combined influence of two factors, frequency of occurrence and provision of word meaning, under the following two conditions, which are characteristic of many real-life reading situations: (a) Advanced L2 learners read an L2 text with the purpose of gaining an overall understanding of it; (b) The text contained a number of words with which they were not familiar.

The L2 learners in this study read a text under one of three conditions: Marginal Glosses (MG), Dictionary (D), and Control (C). For students in the MG group, an L1 translation was provided in the margin for the targeted, unfamiliar words. Students in the Dictionary group were free to use a dictionary. Students in the Control group were not given marginal glosses, nor did they have a dictionary at their disposal.

We entertained the following hypotheses:

Hypothesis 1. Retention of the meaning of unfamiliar target words will be higher in the MG than in the D condition. Retention will be lowest in the C condition. This hypothesis will hold for retention of both the exact and approximate, partial meanings of the target words. Hypothesis 1 is grounded on two considerations: (a) Glosses are provided for *all* target words in the MG condition, whereas the dictionary is likely to be used only for some but not all target words in the D condition; (b) In the C condition, students are likely to ignore some target words or unsuccessfully infer their meanings.

Hypothesis 2. As for the retention of *exact* meanings, an interaction is expected to be found between the Frequency and the Group factor: MG students will profit most from reoccurrence; C students will profit least; D students will profit moderately. Hypothesis 2 is grounded on the following reasoning. Generally, reappearance of a word form will enhance incidental learning of the word form. Furthermore, word reappearance may enhance the learning of the form-meaning connection because each time the word reoccurs, the provided (MG), looked up (D), or inferred (C) form-meaning relation will be reinforced. However, if the exact meaning of a word cannot be inferred from any of the contexts in which the word reoccurs, it cannot be stored in the mental lexicon. Obviously, nonstored information cannot be retained. Thus, frequency of occurrence will positively affect incidental learning of the *exact* word meaning only for students in the MG group or for those students in the D group who do look up the exact meaning.

METHOD

Design

Advanced learners of French as an L2 read a short story by Guy de Maupassant under one of three conditions: Marginal Glosses (MG), Dictionary Use (D), or Control (C), the C group not receiving additional information. Students were instructed to read the text and prepare to answer comprehension questions. Two groups of eight words, each occurring once or three times in the text, were selected as targets. After reading the text, students were tested for their receptive knowledge of the 16 target words.[2]

Subjects

The subjects participating in this experiment were 78 Dutch first-year university students of French from three Dutch universities: the Vrije Universiteit of Amsterdam ($N = 18$), the University of Amsterdam ($N = 34$), and the University of Nijmegen ($N = 26$). These students were considered to be advanced learners of French because they had studied French for 6 years in high school and almost one year at the university.

Reading Materials

The text that students read was "Menuet," a short story by Guy de Maupassant. It relates the story of the narrator who, during one of his early morning strolls in the tree nursery of the Jardin du Luxembourg in Paris, meets an old man. He discovers that the old man had been a leading dancer at the royal opera house under King Louis XV. The old dancer promises to bring his wife, the once famous Castris, during his next walk that afternoon. The writer meets the dancing couple later that day, and they perform a menuet as it was danced in their prerevolutionary days of glory, with the deeply moved writer as the only spectator.

The text contains many low frequency words. Sixteen of these, estimated to be unfamiliar to most first-year students, were selected as target words. These estimations were based on extensive piloting. The text was slightly adapted in

the following way. The first paragraph, a lengthy philosophical introduction that is not crucial to the understanding of the story itself, was left out altogether. In addition, a few phrases were altered so that 8 of the 16 target words appeared three times (F3 words) and 8 appeared once (F1 words). The adapted text contained 1306 words (three pages). Pilot testing had shown that it was extremely difficult to infer the *exact* meaning of the 16 target words from context. In fact, for most of these words, the reader was able to infer only an *approximate* meaning (e.g., "something negative" for *falot* in *spectres falots,* or "a kind of moving or dancing" for *gambader*).

For students in the MG group, the text was further adapted in the following way. The 16 target words, along with 16 other low frequency words, were printed in bold face, and a Dutch translation was printed in the right-hand margin. The words occurring more than once were glossed only on their first occurrence. Marginal glosses for the 16 additional low frequency words were provided in an attempt to render the marginal glosses for the 16 targeted words less salient. The target words consisted of six verbs, four nouns, five adjectives, and one adverb. Table 1 shows the target words by frequency of occurrence.

Testing Materials

Three posttests were administered. The first posttest was a combined recognition and recall test. It contained a list of 32 words, consisting of the 16 target words and 16 words that had not appeared in the text (taken from another Maupassant story). The part of speech of each word was indicated in parentheses (V for verbs, N for nouns, and A for adjectives and adverbs). Students were asked to mark whether the word had appeared in the text (yes/no) and to write down its meaning (in Dutch or French). Students in the Dictionary group were also asked to indicate whether they had looked up the word during the previous reading session (yes/no).

The second test aimed to measure preknowledge of the target words. Students were shown each of the 16 target words with their part-of-speech label in parentheses. They were asked whether they had been familiar with each word before reading the text (previously known/previously unknown).

In the third and final posttest, students were asked to once again provide the meanings of the 16 target words. This time, the target words were not given in isolation (as in Test 1), but in the context of a few lines taken from the original text.

Procedure

Four reading plus testing sessions were held: Session 1 with 18 students of the Vrije Universiteit, Session 2 with 26 students of the University of Nijmegen, Session 3 with 16 students of the University of Amsterdam, and Session 4 with another 18 students of the University of Amsterdam. In each session, students were randomly assigned to the MG, D, and C conditions. They were given the following information: (a) Students would first read a text and then answer comprehension questions; (b) Students would not have the text at their disposal when answering the comprehension questions; (c) Twenty-five minutes would be allotted for reading the text; (d) Some students would be allowed to use a dictionary. (Students in the D group were provided with a French-Dutch dictionary.)

Thus, students were not told in advance that

TABLE 1
The 16 Target Words by Frequency (F1 and F3)

F1 words		F3 words	
braillard (A)	noisy	*charmille* (N)	hornbeam
délectation (N)	delight	*désemparer* (V)	to cease
s'embrouiller (V)	to get confused	*fluet* (A)	thin
falot (A)	pallid	*gambader* (V)	to gambol
frétiller (V)	to tremble	*inopinément* (A)	unexpectedly
harceler (V)	to haunt	*pépinière* (N)	tree nursery
ineffable (A)	ineffable	*simagrée* (N)	gramace
tortiller (V)	to twist	*touffu* (A)	thickly grown

Note. Subjects in the study were native speakers of Dutch learning French. None of the target French words had cognate equivalents in Dutch.
N = noun, V = Verb, A = adjective/adverb.

they would later be tested on their knowledge of the meanings of words in the text; instead, they were told that they would have to answer comprehension questions after reading. This was done to create conditions conducive to incidental vocabulary learning: Students' attention was turned away from *particular unknown words* and directed towards an understanding of the *text as a whole.* Thus, instead of having to answer comprehension questions for which they had prepared, students were tested on their knowledge of vocabulary. The administration of the three posttests took about 30 minutes. The entire session (reading plus posttests) lasted 60 minutes.

RESULTS

Scoring Procedures

The responses of all tests were coded by two of the researchers, independently of each other. In a joint session, they first inspected a few students' translations in Tests 1 and 3 and created a list of responses to be credited with a whole point (fully correct) or a half point (partly correct). They then independently judged all responses of all students. Interrater reliability was more than 99%. The few remaining differences were then resolved in a final joint session.

Recognition

The recognition task, which formed a part of Test 1, had *not* been included to test our main research questions, but rather to check whether frequency of occurrence had played a role at the superficial level of recognition. If this had not been the case, it would have been extremely unlikely to find *any* frequency effect at the level of recall. Thus, the main purpose of the recognition measure was to check whether a basic methodological requirement had been met. Table 2 indicates that this had indeed been the case.[3] A 2 (Frequency) x 3 (Group) MANOVA with repeated measures on the Frequency factor yielded a significant Frequency effect [$F(2,75) = 44.72$, $p < .01$]. There was no significant Group effect, and the Frequency x Group interaction was not significant. Thus, as expected, students, equally in all three groups, more readily recognized target words that had appeared three times than target words that had appeared only once.

TABLE 2
Number of Target Words with Frequencies 1 and 3, Correctly Recognized in Test 1 as Having Appeared in the Text (Max = 8)

	F1 Words		F3 Words	
	M	*SD*	*M*	*SD*
Marginal Glosses group	4.3	1.8	5.8	1.6
Dictionary group	4.3	2.1	5.8	1.5
Control group	4.7	1.6	6.2	1.2

Preknowledge

On the basis of the pilot study, we had expected students to be unfamiliar with the target words. Furthermore, we had anticipated that the few words reported in Test 2 as familiar would be almost equally divided over frequency classes and subject groups. Fortunately, these expectations were realized. Reported preknowledge raw scores, elicited in Test 2, were corrected for translation performance, elicited in Test 1, as follows: If students reported in Test 2 that they had been familiar with a target word before the reading session, but if they had not provided a correct translation in Test 1, the reported preknowledge was not acknowledged. This often happened with the target words *désemparer, s'embrouiller, ineffable,* and *inopinément.* Several students reported in Test 2 that they had known these words before the reading session but failed to give a correct translation in Test 1. Their responses in Test 1 showed that they had analyzed the morphological shape of these words incorrectly and hence had confused these words with, respectively, *s'emparer, être brouillé avec quelqu'un, ineffaçable,* and *opinion.*

Reported preknowledge, corrected for Test 1 performance, was almost nonexistent, as shown in Table 3. The overall mean score is 0.2 out of 8 words. Thus, we had been successful in selecting target words with which students had not been familiar.

TABLE 3
Preknowledge: Average Numbers of Target Words, with Frequencies 1 and 3, with which Students Were Already Familiar (Max = 8)

	F1 Words	F3 Words
Marginal Glosses group	.4	.3
Dictionary group	.0	.1
Control group	.3	.1

Dictionary Use

In Test 1, the 24 students in the Dictionary group were requested to indicate which target

words they had looked up in the dictionary during the reading session. Average reported dictionary use was very low indeed. This was confirmed by the researchers who had been present during the entire session and had particularly observed students in the dictionary group. Of the 24 students, 20 used the dictionary at least once.[4] These 20 students looked up only 38 target words, which is an average of 1.9 out of 16 target words per subject, or 12%.[5]

An interesting pattern developed: Of the 38 words looked up, 34 had occurred three times. The word most often looked up was *pépinière*, the tree nursery in the Jardin du Luxembourg where the meetings took place. Note that for a global understanding of the text, it was not necessary to know the exact meaning of *pépinière*. One could infer from context that this word referred to an area within the Luxembourg gardens.[6]

Incidental Learning: Knowledge of Target Words in Isolation

The responses of Test 1, in which students provided meanings of the target words in isolation, that is, without the context in which they had appeared in the text, were credited with a whole point if completely correct, a half point if partly correct (e.g., "a kind of moving or dancing" for *gambader*), and a zero if completely incorrect or if no response was given. However, whole and half points were not credited when students had reported to know these words beforehand. Thus, Test 1 scores were corrected for Test 2 scores (preknowledge) in order to arrive at the most conservative retention estimate possible in this study. Table 4 shows the average scores for full points only as well as full plus half points.

Both sets of scores were submitted to MANOVA, with Group as the between-subject factor (MG, D, and C) and Frequency as the within-subject factor (F1 and F3). The MANOVA on the fully plus partially correct responses yielded a significant Frequency effect [F (1,75) = 48.49, $p < .001$] and a significant Group effect [F (2,75) = 13.61, $p < .001$], but no significant Frequency x Group interaction. Posthoc multiple-range tests (Scheffé) between groups, on F1 and F3 words separately, revealed that performance of the D and C groups did not differ significantly for either F1 or F3 words, but that the MG group always performed higher than each of the other groups.[7]

The MANOVA on only the full-point scores yielded a significant effect for Frequency [F (1,75) = 34.24, $p < .001$] and a significant effect for Group [F (2,75) = 15.14, $p < .001$]. In this analysis, the Group x Frequency interaction was significant [F (2,75) = 5.80, $p < 01$] as predicted. Posthoc Scheffé tests between groups, on F1 and F3 words separately, revealed that performance of the D and C groups did not differ significantly for F1 or F3 words, but that the MG group always performed higher than each of the other groups.

Detailed analyses per item and student revealed that the retention of the 38 words looked up by students in the Dictionary group was 22.5 (20 fully correct and 5 partially correct responses). This is a retention score of 59%. Even the retention score of the word *pépinière* was 15 out of 17 (88%). Retention of *pépinière* in the MG group was lower: 19 (17 fully and 4 partially correct responses) among 27 students (70%).

Hypothesis 1 predicted highest retention scores for the MG group, lowest for the C group, and scores in between for the D group. This prediction was partly supported. Overall, the MG group had the highest scores in all four cases (fully correct scores only and fully plus partially correct scores, on F1 and F3 words), and, overall, the D group did not perform significantly higher than the C group. The latter result can be simply explained by the fact that students in the D group seldom used the dictionary. They thus processed the words not looked up in the same manner as students in the C group, hence their retention scores were as low.

TABLE 4
Retention Performance, Corrected for Preknowledge, of the Meanings of Target Words with Frequencies 1 and 3, Presented without Context in Test 1 (Max = 8)

	Full Points Only				Full Plus Partial Points			
	F1 Words		F3 Words		F1 Words		F3 Words	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Marginal Glosses group	1.3	2.0	2.6	1.9	1.4	2.0	2.8	2.0
Dictionary group	0.2	0.4	1.0	0.8	0.2	0.4	1.2	0.9
Control group	0.3	0.6	0.6	0.6	0.4	0.6	1.0	0.9

However, in the few cases that D students did use the dictionary, their retention scores were even higher than those of the MG group.

Hypothesis 2 predicted a significant interaction between the Frequency and Group factors on the exact meaning scores. This hypothesis was supported. As has already been reported, the Frequency x Group interaction was significant.

Incidental Learning: Knowledge of Target Words in Context

Table 5 gives means and standard deviations of performance in Test 3, in which word knowledge was tested with the help of contextual information. Because the contexts in Test 3 provided a cue for the correct responses, whereas no cues had been given in Test 1, mean retention scores in Test 3 were higher than in Test 1 (compare Table 5 with Table 4).

The scores of Test 3 were submitted to the same analyses as those in Test 1. The MANOVA on fully and partially correct scores yielded a significant Frequency effect [F (1,75) = 218.11, $p < .001$] and a significant Group effect [F (2,75) = 17.47, $p < .001$], whereas the Frequency x Group interaction almost reached the .05 probability value [F (2,75) = 3.04, $p = .054$]. Posthoc Scheffé tests revealed that performance of the D and C groups did not differ significantly on F1 or on F3 words, whereas the MG group always significantly outperformed each of the two other groups.[8]

The MANOVA on only the fully correct scores yielded a significant Frequency effect [F (1,75) = 106, $p < .001$], a significant Group effect [F (2,75) = 20.01, $p < .001$], and a significant Frequency x Group interaction [F (2,75) = 8.29, $p < .01$]. Posthoc Scheffé tests yielded the same pattern as before: The D and C groups did not differ significantly on F1 or on F3 words, whereas the MG group performed significantly better than each of the other groups on both F1 and F3 words.

Detailed analyses on the 38 words looked up by students in the Dictionary group yielded a result almost identical to that in Test 1. Retention was 23 (20 fully correct and 6 partially correct responses) compared to 22.5 in Test 1. Thus, provision of contextual cues in Test 3 did not lead to a higher retention of words looked up during reading.

These Test 3 results yielded the same pattern as the results of Test 1. Again, there is partial support for Hypothesis 1, that is, the MG group had higher scores than the other groups, as predicted. However, because the D group students used the dictionary so infrequently, their overall retention scores were as low as those of the C group students, whereas their retention scores of the words that they did look up were somewhat higher than even those of the MG group students. As predicted in Hypothesis 2, the Frequency x Group interaction was significant in the analysis on only fully correct responses, but not in the analysis on fully plus partially correct responses.

Some students in the MG group demonstrated that it was possible to learn many words from the glosses. From the 27 students in this group, 4 had full or partial knowledge of 13 target words or more (with preknowledge subtracted from their Test 3 score).

A scrutiny of individual responses revealed that none of the students in the Control and Dictionary groups were able to provide a fully correct response to four target words, unless these words were previously known or looked up in the dictionary. These were two F3 words, *pépinière* and *charmille,* and two F1 words, *s'embrouiller* and *falot.* The MG group ($N = 27$) demonstrated that both the provision of marginal glosses and the reoccurrence of words in the text facilitated vocabulary learning with respect to these four words. Fully correct responses in Test 3 were given by 20 and 10 students for the two F3 words, *pépinière* and *charmille* respectively, and by 4 and 6 students for the two F1 words, *s'embrouiller* and *falot* respectively. It is interesting

TABLE 5
Retention Performance, Corrected for Preknowledge, of the Meanings of Target Words with Frequencies 1 and 3, Presented in Context in Test 3 (Max = 8)

	Full Points Only				Full Plus Partial Points			
	F1 Words		F3 Words		F1 Words		F3 Words	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Marginal Glosses group	2.9	1.7	3.9	2.2	3.4	1.6	4.4	2.0
Dictionary group	1.1	1.1	1.8	1.0	1.5	0.9	2.5	1.0
Control group	1.3	1.0	1.3	0.9	1.8	0.9	2.4	0.9

to note that students in the Dictionary group looked up only *pépinière* and *charmille,* the two F3 words, and not *falot* or *s'embrouiller,* the two F1 words.

Check on the Influence of University Enrollment

Because students' knowledge of French vocabulary differed between universities in the pilot study, we took measures in the present study to ensure that the university factor did not interfere with the group factor by randomly assigning students of each university to the three conditions. In order to verify that the university factor had not interfered with the experimental design, we conducted one way ANOVAs on all variables (preknowledge, retention of isolated words, and retention of words in context, for F1 and F3 words separately), with university as the independent factor. There were 18, 26, and 34 students respectively from the Vrije Universiteit of Amsterdam, the University of Nijmegen, and the University of Amsterdam. In none of these six analyses was a main university effect found. Thus, in the present study, we were successful in eliminating the potentially confounding influence of university enrollment.

SUMMARY

Before discussing the results, let us summarize the main findings of this study (N = 78) and, to the extent that reliable data are available, of the pilot study (N = 63).

1. There was substantial incidental vocabulary learning in its most modest form. The recognition of word forms and reappearance of words had clear effects on word recognition: F1 and F3 words were correctly recognized in four and six out of eight words respectively (see Table 2).

2. Overall, the provision of marginal glosses resulted in much better retention scores than the provision of dictionaries. The MG group performed twice as well as or better than the D group on F3 and F1 words in Test 1 and 3 respectively (see Table 6). This evidence should be given more weight in the case of Test 1 than in the case of Test 3 because Test 1 provided a measure of information processing during *reading,* whereas performance on Test 3 may have stemmed from inferencing during *test taking.*

3. Students in the D group seldom used their dictionary. On average, they looked up only 12% and 15% of the target words in the main and pilot study respectively.

4. However, when students in the D group did look up a word, their chance of remembering its meaning was greater than the average retention in the MG group (see Table 6, rows 1 and 3).

5. As predicted by Hypothesis 2, frequency of occurrence of the target words in the text had a significant impact on the retention of exact word meanings, but not on the retention of exact plus partial word meanings. It was predicted that if the exact word meaning could not be inferred with certainty from any of the three contexts in which F3 words appeared, reoccurrence itself could only affect the retention of partial meanings. This was indeed the case. However, since the exact meanings were known with certainty only to students in the MG group and to those students in the D group who had actually consulted the dictionary, it was only these students who were expected to gain from the reoccurrence of target words, which indeed they did.

TABLE 6
Retention Performance of Target Words with Frequencies 1 and 3, in Test 1 (Presented in Isolation), in Percentages (Corrected for Preknowledge, Full Plus Partial Points)

	F1 Words	F3 Words
Marginal Glosses group: overall[a]	18	35
Dictionary group: overall[b]	3	15
Dictionary group: only words looked up[c]	25	63

[a] Figures calculated from Table 4, line 1: 1.4/8 and 2.8/8 respectively.
[b] Figures calculated from Table 4, line 2: 0.2/8 and 1.2/8 respectively.
[c] Respectively 1 out of 4 F1 words and 21.5 out of 34 F3 words looked up.

DISCUSSION

Relevance as a Mediating Variable

As was mentioned in the introduction, previous research has shown support for the claim that "meaning inferred" yields higher retention than "meaning given" (Hulstijn, 1992; Mondria & Wit-de Boer, 1991; Watanabe, 1992). In the present study, better retention was attained by the MG than by the C group. Is this result at variance with the earlier finding or can the two findings be reconciled? We believe that there is no inconsistency and that all findings can be interpreted harmoniously. In the studies of

Hulstijn (1992), Mondria and Wit-de Boer (1991), and Watanabe (1992), measures were taken to guarantee that students in the Meaning Inferred conditions did indeed attempt to infer meanings of the unfamiliar words. In the present study, however, C and D group students simply ignored many of the target words because, apparently, students did not deem their meaning essential to an understanding of the text as a whole. This posthoc interpretation of the C and D groups' reading "behavior" (an appropriate term even in a cognitive, post-behavioristic era) is supported by the low incidence of dictionary consultations in the D group. Students in the D group did not consider it necessary to look up the meanings of most target words: *not,* however, because they had successfully inferred word meanings without the help of the dictionary! On the contrary, their low retention scores in Test 3, in which the target words were presented in their context, clearly show that they had not. The true reason that students did not look up these words may be that they did not perceive them as relevant in the context of their reading goal. What Miller and Gildea (1987) observed for fourth-grade children reading L1 texts also appears to apply to advanced L2 readers: "One trouble with this approach is that most healthy, right-minded children have a strong aversion to dictionaries" (p. 89). The conclusion to draw from this is that advanced L2 learners, when reading a text for global comprehension, do not feel an urge to interrupt the flow of reading by investing considerable time and mental effort to infer or look up the meaning of unknown words. They will only do so when they perceive the word as relevant, when intrigued by it, or when alerted or even annoyed by the fact that the word keeps returning in the text and may warrant their attention.

Further support for this interpretation is the case of *pépinière,* the word most frequently looked up by the D group and best remembered by the MG group. Why look up *pépinière* and not any of the other seven words which occurred three times in the text? We believe that this word was perceived as more relevant than the other words because it referred to the location where the meetings between the narrator and the elderly dance couple took place. Further credence to this interpretation accrues from the following finding: The range of the 34 F3-word consultations by D group students varied from 1 to 17 per word, with *pépinière* at the top of the list with 17 consultations. The rank order based on these F3-word consultations in the D group corresponded fairly well (with one exception) to the rank order of retention scores of these F3 words in Test 1 by the MG group. This leads us to speculate that if dictionary consultations were motivated by perceived word relevance, retention of word meanings provided in the margin may also have been influenced by perceived relevance.

It would be worthwhile to investigate, in future research, whether perceived (ir)relevance of unknown words interacts with text genre. Perhaps readers perceive unknown words as less relevant, in general, when reading a fictional text than when reading an expository text. Often, propositions in a fictional text are structured according to the Collection principle, whereas the relationship between propositions in an expository text is often of the Adversative, Explanation, Consequent, and Antecedent types (Bossers, 1992; Meyer, 1975; Meyer & Rice, 1984). This structural difference between fictional and expository texts may elicit more shallow and deep ways of word processing on the part of the reader.

The Probabilistic Nature of Inferencing: Good and Poor Guesses

In the present study, a broad distinction was made between two possible outcomes of inferring activities: exact versus approximative meanings. These two outcomes were linked with fully and partially correct responses on the retention tests. The relation, however, is not straightforward. Future research should pay more attention to this issue, which is complicated by several causes. First, the outcome of the inferring activity is influenced by at least two factors, (a) the strength of the contextual cues (a text factor) and (b) the reader's verbal and nonverbal proficiency (a reader factor). In addition, the textual factor cannot always be treated as orthogonal to the reader factor because the strength and comprehensibility of the contextual cues may depend on the reader's proficiency.

The second cause of the issue's complexity is the probabilistic nature of inferences and the fact that, in principle at least, the probability to infer semantic features may not be the same for all features alike. Let us take the word *gambader* for instance. It was not difficult to infer that it meant some sort of motion, but it was almost impossible to infer that it meant "to gambol" ("to make jumping movements"; the Dutch

equivalent is *huppelen*.). Similarly, it was virtually impossible to infer the precise meaning of *charmille* because it could be any kind of tree in a tree nursery. If any subject in the C group had given the correct response ("hornbeam"), it could only have been the result of purely random guessing.

To our knowledge, research on contextual vocabulary learning, including the present study, has not sufficiently distinguished these complicating factors.[9] In future research, it should be possible to bring the differential effects of "meaning given" and "meaning inferred" to the forefront by categorizing the contextual information of a to-be-inferred target word into two elements, according to the following two questions: (a) Which part of the word's meaning can be "inferred" with certainty? and (b) Which part of the word's meaning can only be "guessed" with some degree of plausibility but not with certainty? This categorization should allow researchers to make the following, more principled distinctions: (a) to the extent that a completely correct inference can be made on the basis of the contextual information, a distinction between a correct and an incorrect inference; (b) to the extent that a completely correct inference cannot be made on the basis of the contextual information, a distinction between a well-motivated but random guess, on the one hand, and a poorly motivated, wild guess, on the other (see Table 7).

CONCLUSIONS

The present study aimed to explore how the generally low incidence of incidental vocabulary learning can be improved. This exploration was accomplished by investigating the combined influence of two factors that previous research had shown to be conducive to incidental learning: (a) frequency of occurrence and (b) provision of word meaning (through marginal glosses or dictionary use). Support was found for the hypothesis that frequency of occurrence will foster incidental vocabulary learning more when advanced L2 readers are given the meanings of unknown words through marginal glosses or when they look up these meanings in a dictionary than when no external information concerning the meanings of unknown words is available. In the former case, reappearance of a word will reinforce the form-meaning connection in the reader's mental lexicon. In the latter case, readers will often ignore unknown words or unsuccessfully infer their meanings, which will limit the frequency effect. Furthermore, support was found for the hypothesis that, generally, the effect of marginal glosses will be greater than that of dictionary use because readers often do not make use of the dictionary. However, when readers *do* use the dictionary, the incidence of incidental vocabulary learning will be as good as, or even better than, when they are provided with marginal glosses.

In addition, the present study, along with our earlier research (Hulstijn, 1992), suggests that intermediate and advanced L2 learners, when reading a fictional or nonfictional text of more than one page in order to understand the main idea, seldom use a dictionary. Only when they read a short text (Knight, 1994) or when dictionary consultation has been made extremely easy by a simple click of the mouse in a computer presentation (Hulstijn, 1993), do some L2 readers look up the meanings of unknown words. Learners tend to ignore unfamiliar words, except when they perceive the words as relevant for reaching their reading goal or when they notice that an unknown word keeps reappearing and might therefore be worth their attention.

When we consider the height of the retention scores, we draw two complementary conclusions. On the one hand, this study, along with the pilot study which preceded it, has shown that even when readers consult the dictionary,

TABLE 7
Categories of Response Quality Assessment

Can feature be inferred completely?			
YES		NO	
Good response	*Poor response*	*Good response*	*Poor response*
Correct inference	Incorrect inference	Well-motivated guess	Wild but correct guess
			Incorrect and poor guess

when the meanings of unfamiliar words have been made available through marginal glosses, or when such words appear not once but three times in the text, readers usually forget almost immediately the meanings of more than half of the words thus processed. On the other hand, this study, along with the studies mentioned in the introduction, has demonstrated that incidental vocabulary learning during reading does occur and that it has the potential of contributing substantially to an incremental process of vocabulary acquisition (Nagy, Herman, & Anderson, 1985).

PEDAGOGICAL RECOMMENDATIONS

Intermediate and advanced L2 learners enlarge their vocabulary to a great extent through incidental learning during reading activities. There is no doubt that extensive reading is conducive to vocabulary enlargement. However, reading for global meaning alone will not do the job. For words to be learned, incidentally as well as intentionally, learners must pay attention to their form-meaning relationships. Learners should therefore be encouraged to engage in elaborating activities, such as paying attention to unfamiliar words deemed to be important, trying to infer their meanings, looking up their meanings, marking them or writing them down, and reviewing them regularly. There is empirical evidence that good L2 learners actively manage their vocabulary learning and regularly review their records of new words (see Sanaoui, 1995, and research reviewed there).

Teachers and material developers can foster vocabulary learning through reading in various ways (see also Hulstijn, in press):

1. First and foremost, assign learners reading texts that are interesting and motivating. If the text does not alert their curiosity, learners will not be willing to devote the required mental effort to unfamiliar words.

2. Facilitate the burden of dictionary use because dictionary use interferes with the process of constructing a mental representation of text meaning. Provide marginal glosses when the text is read on paper, or provide easy-to-access electronic glosses when the text is read on computer. For those words whose meaning can indeed be inferred completely from the context, provide a symbol (e.g., an asterisk) so that readers know that they should first try to infer the meaning before consulting the gloss or dictionary.

3. Make important target words reappear several times, if possible. (This is not possible, of course, when learners read an unaltered, original text.)

4. Give learners a list of important words for subsequent intentional learning, or, perhaps more motivating, encourage learners to draw up an individual list of words that they consider relevant to remember.

5. Invite learners to review regularly these word lists. To that end, add vocabulary exercises to the printed text, or incorporate such exercises into the computer courseware.[10]

What the guidelines above intend to bring about is *not to replace* incidental learning by intentional learning, but *to follow up* on incidental learning with intentional learning. Since intentional learning may require considerable effort, time, and, hence, motivation on the part of L2 learners, it is all the more important that teachers and course designers provide materials that the learners consider relevant and interesting. In the words of Haastrup (1989): "Learners will always find out the meaning of words that are important to them" (p.43). If a desire to know a word's meaning is the driving force, the mental elaboration required for acquisition will then come of its own accord. However, it is the teacher's and course designer's task to provide learners with easily accessible glosses and learner-friendly, nontedious review opportunities.

NOTES

[1] We would like to thank the following colleagues for providing us with testing opportunities at the French departments of the Universities of Nijmegen, Utrecht, and Amsterdam and of the Vrije Universiteit of Amsterdam: Chris de Kok, Lydius Nienhuis, Elisabeth van der Linden, and Marion Suttorp (respectively). We are also indebted to Bart Bossers (Vrije Universiteit) and four anonymous reviewers for their thorough comments on an earlier version of this text.

[2] In fact, not one but two studies were conducted: the study reported in this paper and a pilot study that was conducted a year earlier. The pilot study used the same Guy de Maupassant text as well as first-year university students of French ($N = 63$). The pilot study, however, had several methodological shortcomings and so is not extensively reported in this paper. The main differences between the pilot study and the one reported here are as follows: Students in the pilot study were not randomly assigned to reading conditions; there were three classes of students (at the Vrije Universiteit of Amsterdam, the University of Amster-

dam, and the University of Utrecht) and conditions were assigned to entire classes. We had not expected the three classes to differ substantially in French proficiency. However, results indicated that target-word preknowledge in the D group was substantially and significantly lower than in the other two groups. This fact interfered with the reading condition factor (MG, D, and C). In the pilot study, we had selected 15 target words, three groups of 5 words each, which occurred in the text once, twice, or three times (F1, F2, F3). We had anticipated that only very few students would be familiar with these words and that preknowledge, if existent at all, would be evenly distributed among the three frequency classes. This, however, turned out not to be the case. As it happened, F2 words were significantly more familiar than F1 and F3 words. Thus, preknowledge interfered with the frequency factor as well. In the remainder of this paper (i.e., in notes 3 through 8), we will report on the pilot study results only when the data were reliable. That is, no results will be reported of the retention of F2 words in comparison to F1 and F3 words, or of the D group in comparison with the MG and C groups.

[3] A similar result was found in the pilot study: F3 words were significantly more often recognized than F2 words, and F2 words significantly more often than F1 words. There was no significant Group effect (MG, D, and C).

[4] In the pilot study, a similar result was obtained: 18 of 22 students had used the dictionary at least once.

[5] In the pilot test, the incidence of dictionary use was at a low level as well: The average was 2.3 look-ups out of 15 target words (15%).

[6] In the pilot study, the words most frequently looked up were *pépinière* and *charmille*. As in the main study, these words, whose precise meanings could not be inferred from the context, appeared three times in the text.

[7] In Test 1 of the pilot study, students in the MG group remembered F3 words ($M = 1.95$ out of a maximum score of 5) significantly better than students in the C group ($M = .95$), whereas no significant differences in retention of F1 words were found. This finding lends support to the first leg of our interaction hypothesis. Because, as explained in Note 2, students in the Dictionary group happened to have more preknowledge than those in the other groups, the second leg of the hypothesis could not be tested.

[8] In Test 3 of the pilot study, students in the MG group significantly outperformed students in the C group on both F1 and F3 words. Out of a maximum score of 5, the mean scores (corrected for preknowledge) for the MG and C group were, respectively, 2.45 versus 1.57 for F1 words, and 2.35 versus 1.48 for F3 words. Thus, the MG group performed better, but no frequency effect was obtained. (Due to the methodological problems mentioned in Note 2, the data of the F2 words and the data of the Dictionary group were not reliable and were therefore omitted in the analysis.)

[9] However, in the study by Hulstijn (1993), the Inferability factor was experimentally manipulated, distinguishing target words of high or low inferability.

[10] In a recent investigation using the same French text and the same type of L2 learners (Dutch first-year university students of French) as in the present study, we compared two conditions. For one student group, all target words appeared once in the text with glosses in the margin. This group was required to do a simple vocabulary reviewing exercise, consisting of fill-in-the-blank questions, immediately after reading. For the other student group, target words appeared five times in the text, the first time glossed. This group did not do a reviewing exercise after reading. Retention of word meanings on the immediate posttests (words in isolation and in context) was substantially and significantly higher for the one-exposure-plus-exercise group than for the five-exposures-without-exercise group.

REFERENCES

Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading, 7,* 15–32.

Bossers, B. (1992). *Reading in two languages: A study of reading comprehension in Dutch as a second language and in Turkish as a first language.* Unpublished doctoral dissertation, Vrije Universiteit Amsterdam.

Brown, H. D. (1994). *Teaching by principles: An interactive approach to language pedagogy.* Englewood Cliffs, NJ: Prentice Hall Regents.

Carnine D., Kameenui, E. J., & Coyle, G. (1984). Utilization of contextual information in determining the meaning of unfamiliar words. *Reading Research Quarterly, 19,* 188–204.

Coady, J. (1993). Research on ESL/EFL vocabulary acquisition: Putting it in context. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 3–23). Norwood, NJ: Ablex.

Davies, J. N. (1989). Facilitating effects of marginal glosses on foreign language reading. *Modern Language Journal, 73,* 41–48.

Dubin, F., & Olshtain, E. (1993). Predicting word meanings from contextual clues: Evidence from L1 readers. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 181–202). Norwood, NJ: Ablex.

Ellis, R. (1994). Factors in the incidental acquisition of second language vocabulary from oral input: A review essay. *Applied Language Learning, 5,* 1–32.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly, 25,* 375–406.

Haastrup, K. (1989). The learner as word processor. *AILA Review, 6,* 34–46.

Hazenberg, S., & Hulstijn, J. H. (1996). Defining a

minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics, 17,* 145–163.

Huckin, T., & Haynes, M. (1993). Summary and future directions. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 289–298). Norwood, NJ: Ablex.

Hulstijn, J. H. (1989). Implicit and incidental second language learning: Experiments in the processing of natural and partly artificial input. In H.W. Dechert & M. Raupach (Eds.), *Interlingual processes* (pp. 49–73). Tübingen, Germany: Gunter Narr.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P.J.L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 113–125). London: Macmillan.

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

Hulstijn, J. H. (in press). Mnemonic methods in foreign-language vocabulary learning: Theoretical considerations and pedagogical implications. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy.* New York: Cambridge University Press.

Jacobs, G. M., Dufon, P., & Fong, C. H. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading, 17,* 19–28.

Jones, F. R. (1995). Learning an alien lexicon: A teach-yourself case study. *Second Language Research, 11,* 95–111.

Knight, S. (1994). Dictionary: The tool of last resort in foreign language reading? A new perspective. *Modern Language Journal, 78,* 285–299.

Krantz, G. (1991). *Learning Vocabulary in a Foreign Language: A Study of Reading Strategies.* Göteborg, Sweden: Acta Universitatis Gothburgensis.

Laufer, B., & Sim, D. D. (1985). Taking the easy way out: Non-use and misuse of clues in EFL reading. *English Teaching Forum,* 7–10.

Meyer, B. J. F. (1975). *The organization of prose and its effect on memory.* Amsterdam: North Holland.

Meyer, B. J. F., & Rice, G. E. (1984). The structure of text. In P.D. Pearson (Ed.), *Handbook of reading research.* New York: Longman.

Miller, G., & Gildea, P. M. (1987). How children learn words. *Scientific American, 257 (2),* 86–91.

Mondria, J- A., & Wit-de Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics, 12,* 249–267.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24,* 237–270.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20,* 233–253.

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* New York: Newbury House.

Nation, I. S. P., & Coady, J. (1988). Vocabulary and reading. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 97–110). London: Longman.

Oskarsson, M. (1975). On the role of the mother tongue in learning foreign language vocabulary: An empirical investigation. *ITL Review of Applied Linguistics, 27,* 19–32.

Sanaoui, R. (1995). Adult learners' approaches to learning vocabulary in second languages. *Modern Language Review, 79,* 15–28.

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System, 6,* 72–78.

Scherfer, P. (1993). Indirect L2-vocabulary learning. *Linguistics, 31,* 1141–1153.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review, 11,* 11–26.

Schouten-van Parreren, C. (1989). Vocabulary learning through reading: Which conditions should be met when presenting words in texts? *AILA Review, 6,* 75–85.

Schouten-van Parreren, C. (1992). Individual differences in vocabulary acquisition: A qualitative experiment in the first phase of secondary education. In P.J.L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 94–101). London: Macmillan.

Summers, D. (1988). The role of dictionaries in language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language learning* (pp. 111–125). London: Longman.

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M.G. McKeown & M.E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89–105). Hillsdale, NJ: Erlbaum.

Swaffar, J. K., Arens, K. M., & Byrnes, H. (1991). *Reading for meaning: An integrated approach to language learning.* Englewood Cliffs, NJ: Prentice Hall.

Watanabe, Y. (1992). *Incidental learning of vocabulary: Retention of inferred meanings vs. given meanings.* Unpublished master's thesis, University of Hawai'i at Manoa.