

THE EFFECTS OF INFERRING, VERIFYING, AND MEMORIZING ON THE RETENTION OF L2 WORD MEANINGS

An Experimental Comparison of the “Meaning-Inferred Method” and the “Meaning-Given Method”

Jan-Arjen Mondria
University of Groningen

This study investigated whether a word-learning method in which learners infer the meaning of unknown words from the context, subsequently verify the meaning with the aid of a word list, and finally memorize the meaning (“meaning-inferred method”) leads to better retention than one in which the meaning of unknown words is given in the form of a translation so that learners can immediately start memorizing (“meaning-given method”). Additionally, the learning effect of the various stages of the meaning-inferred method (inferring, verifying, and memorizing) was investigated. In all cases the amount of time invested was recorded. The most important findings were: (a) The meaning-inferred method leads to a similar level of retention as the meaning-given method, but the former is considerably more time-consuming and therefore less efficient; and (b) each separate stage of the meaning-inferred method leads to retention, but the learning effect of memorizing is the greatest, and the learning effect of verifying is about the same as that of inferring.

This article is based on a study conducted for my doctoral dissertation (Mondria, 1996). I would like to thank Paul Meara for his helpful comments on an earlier version of this article.

Address correspondence to: Jan-Arjen Mondria, University of Groningen, Department of Language and Communication, P.O. Box 716, 9700 AS Groningen, the Netherlands; e-mail: J.A.Mondria@let.rug.nl.

One of the constantly recurring questions with regard to vocabulary acquisition in foreign language teaching is whether learning methods based on inferring the word meaning with the aid of the context lead to better retention than those in which the meaning of a word is “given” (e.g., Haastrup, 1991; Mondria, 1996; Schouten-van Parreren, 1985). On the basis of various studies of first language (L1) and second language (L2) incidental vocabulary acquisition that show that inferring *per se* leads to a certain amount of retention (for L1, see Swanborn & De Gloppe, 1999; for L2, see Huckin & Coady, 1999), it is to be expected that, when a learner infers the meaning of a word before memorization (the “meaning-inferred method”), he or she better retains that meaning than when the meaning of the word is “presented” to the learner—for example, in the form of a translation (the “meaning-given method”). The explanation for the retention effect of inferencing is probably due to deep processing of the unknown word, as a result of which all kinds of links (elaborations) are formed between the word, its meaning, the context, and the already present knowledge of the learner (Anderson, 1990; Ellis, 1995; Hulstijn, 2001). The construct of task-induced involvement, introduced by Laufer and Hulstijn (2001), is a recent attempt to operationalize the construct of elaboration. According to this construct, the cognitive *search* and *evaluation* activities—in addition to the motivational component *need*, which is not relevant here—are conducive to retention. *Search* is defined as the attempt to find the meaning of an unknown L2 word, and *evaluation* is defined as the comparison of a given word with other words or a comparison of a specific meaning of a word with its other meanings.

A precondition for the success of the meaning-inferred method is that the meaning of a word is inferred correctly, as incorrectly inferred meanings will also be retained better. Therefore, it is important that the context in which the unknown word occurs is “pregnant”—that is, it contains sufficient clues for inferring the meaning of the word (Mondria & Wit-de Boer, 1991; Van Parreren, 1967). Additionally, the learners should have sufficient knowledge of the context words and be skillful at inferencing. As all of these preconditions are not always satisfied in practice (see Bensoussan & Laufer, 1984; Haynes, 1993; Kelly, 1990; Laufer, 1997; Laufer & Sim, 1985; Schatz & Baldwin, 1986; Stein, 1993), it is advisable that the learners, after inferring, verify the correctness of the inferred meaning with the aid of a dictionary (Schouten-van Parreren, 1985). Moreover, such a verification stage can contribute to retention in combination with the act of writing down the inferred and verified meaning (Schouten-van Parreren; cf. Thomas & Dieter, 1987). An additional point of interest concerning the meaning-inferred method is the amount of time the procedure takes. As inferring and verifying probably take more time than simply giving the meaning, it cannot be ruled out that the meaning-inferred method, even in the case of a higher learning effect, is still less efficient than the meaning-given method (Kelly; Miller, 1991; Sternberg, 1987).

In the following literature review I examine the empirical evidence for the supposed superiority of the meaning-inferred method over the meaning-given

method. In this review and in the remainder of this article, *incidental vocabulary learning* refers to the learning of vocabulary as the byproduct of any activity not explicitly geared to vocabulary learning. *Intentional vocabulary learning*, on the other hand, refers to any activity aiming at committing lexical information to memory (Hulstijn, 2001; see also Gass, 1999).

LITERATURE REVIEW

Experiments were selected in which the learning effect of meaning-inferred methods and meaning-given methods is compared. Subsequently, all experiments that also differed in other respects than the factor of inferencing were eliminated, as they could not yield a valid conclusion with regard to the research questions of the present study. This concerns Gershman (1970), Jenkins, Matlock, and Slocum (1989), Jenkins, Pany, and Schreck (1978), Johnson and Stratton (1966), Koster (1989), Levine and Reves (1990), McDaniel and Pressley (1984), and Qian (1996), where the meaning-inferred and meaning-given conditions differed from each other with regard to the factor of context (i.e., context vs. no context, a difference in type of context, or a difference in the point in time that the context became available). Fischer (1994) was eliminated because no retention data were reported. Thus, the following 13 studies were left.

Forlano and Hoffman (1937) compared the learning effect of giving the meaning with the learning effect of inferring the meaning followed by feedback within the intentional-learning paradigm. On both the immediate and delayed tests, giving the meaning turned out to lead to significant and substantially higher retention than inferring. A possible explanation is the absence of a context, which may have caused many inferencing errors and thus interfered with the retention of the correct meanings.

Berlyne, Carey, Lazare, Parlow, and Tiberius (1968, exp. 1) showed that, within the incidental-learning paradigm, inferring led to significantly higher retention than giving (immediate and delayed tests). Within the intentional-learning paradigm, on the other hand, inferring was significantly worse than giving (immediate test) or was similar (delayed test). However, the interpretation of these conclusions is complicated by the fact that the analysis was based on the combined results of two different meaning-inferred conditions and two different meaning-given conditions.

Schouten-van Parreren (1980, 1985) compared the learning effect of giving (word meanings in the margin and in a word list) with that of inferring. Students read a text with the aim of text comprehension and vocabulary learning. After the elimination of the youngest group, whose inferencing skills were insufficiently developed, the immediate and delayed isolated-words tests showed no significant differences between giving and inferring. However, on the tests with nonpregnant sentences, the results of the meaning-inferred condition were significantly lower (14% and 8% lower on the immediate and delayed tests, respectively). At the same time, there was a significant ordinal interaction between condition (giving vs. inferring) and time of test (immediate vs. delayed)

on both the isolated-words test and the nonpregnant-sentences test. On the delayed tests, the difference between giving and inferring was smaller than on the immediate tests. This implies that students in the meaning-inferred condition had forgotten fewer words than those in the meaning-given condition.

Bialystok (1983, exp. 2) required students to read stories with the aim of text comprehension and vocabulary acquisition. On the immediate test, retention in the meaning-given condition (with the aid of a dictionary) was significantly higher than in each of the three meaning-inferred conditions. It cannot be ruled out, however, that the lag of the meaning-inferred conditions vis-à-vis the meaning-given condition was caused by incorrect inferencing.

Hulstijn (1989, 1992) carried out five experiments that compared giving and inferring. The most convincing differences on the (more or less immediate) tests are the following. In the incidental-learning circumstances in one experiment (exp. 1), giving (by means of a gloss) was significantly better than each of three forms of inferencing (a text context, which may be supplemented in the margin with a pregnant-sentence context or a multiple-choice gloss), but this difference could not be replicated in a repetition of the experiment under more reliable conditions (exp. 2). However, in another experiment (exp. 5, isolated-words test), inferring (with the aid of a multiple-choice gloss) was significantly better than giving (a synonym in the margin). The fact that inferring was better than giving on the contextual-words test in experiments 3–5 is less convincing because the test contexts were identical to the learning contexts; this makes the test less valid and may have given an advantage to the meaning-inferred conditions. No differences were found in the intentional-learning conditions.

Krantz (1991) required students to read a textbook with the aim of acquiring content and vocabulary knowledge. Unknown words could be looked up in an electronic dictionary. On the immediate test, the meanings of words that had been looked up in a dictionary turned out to be better known than the meanings of words that had not been looked up and therefore were supposed to have been inferred with the aid of the context. However, no reliable conclusion about the superiority of giving or inferring can be drawn, as the category of words that were looked up differs in several respects from the category of words that were supposed to be inferred: text frequency, general frequency, textual context, and possibly importance for the student.

Lupescu and Day (1993), in an incidental-learning experiment, required students to read an L2 text containing words that could be inferred with the aid of the context. On the immediate multiple-choice vocabulary test, students who had a bilingual dictionary at their disposal during reading performed significantly better than students who read without a dictionary. However, as it cannot be ruled out that students tried to infer the meaning before consulting a dictionary, it is not clear whether the conclusion should be that looking up is better than inferencing or that inferencing followed by looking up is better than inferencing. Moreover, it should be pointed out that students who used a dictionary read almost twice as long as those who did not use dictionaries.

Jacobs, Dufon, and Fong (1994) required students to read an L2 text with

or without glosses (incidental learning). Glosses led to a significantly higher score on the immediate vocabulary test. However, on the delayed test there was no longer a significant difference.

Knight (1994) asked students to read two electronic L2 texts (incidental learning). Those who had access to an electronic dictionary during reading performed significantly better on the delayed vocabulary test than those who had to resort to inferencing. However, it cannot be ruled out that looking up sometimes implied inferencing followed by looking up (cf. Luppescu and Day, 1993).

Hulstijn, Hollander, and Greidanus (1996) compared the incidental-learning effect of (a) reading, (b) reading with marginal glosses, and (c) reading with access to a dictionary. On the immediate test, students in the marginal-glosses condition knew more words than students who had to resort to inferencing or to a dictionary. The explanation for the lower performance of the dictionary condition compared to the gloss condition is that students seldom used the dictionary. Incidentally, the retention of the words that were actually looked up was higher than that of the gloss condition.

Watanabe (1997) compared the incidental-learning effect of reading four L2 text versions: (a) original, (b) vocabulary explanations immediately after the word in the text (appositives), (c) single glosses in the margin, and (d) multiple-choice glosses in the margin. The retention of the words in both marginal-gloss conditions was significantly higher than in the other two conditions (immediate and delayed tests). However, there was no significant difference between the “inferencing gloss” (multiple-choice gloss) and the “giving gloss” (single gloss) or between the original text condition (where the unknown word meanings had to be inferred) and the text with appositives (a form of giving).

Grace (1998) had students learn an L2 with the aid of a computer that gave them access to, among other things, pregnant contexts. Students who were given additional access to the translation of these sentences—involving an extra time investment—performed significantly better on the vocabulary tests (immediate and delayed). It is not clear, however, whether this effect was caused by looking up or by the combination of inferring and looking up (cf. Knight, 1994; Luppescu & Day, 1993).

Fraser (1999) found that words whose meanings had been inferred and then looked up had clearly been better retained after 1 week than words whose meanings had only been inferred or only looked up. However, the value of this conclusion is limited by the fact that (a) the words possibly differed by strategy, as the subjects themselves determined the strategy to be used for each word, and (b) the intervening inquiries and retention tests may have affected the incidental-learning character of the experiment. Furthermore, significance data are lacking.

Summary of Literature Review

On the basis of this review and taking into account the limitations of some of the experiments—incomparable words for inferring and giving, artificial forms

of inferencing, interference of intervening inquiries, lack of clarity about the precise activity of the subjects, and absence of a delayed test—the following conclusions emerge for incidental learning.

1. Looking up in a dictionary (giving), possibly after inferring, led to higher retention than inferring (Knight, 1994; Lupescu & Day, 1993).
2. Glosses (giving) led to higher retention than inferencing in two experiments (Hulstijn et al., 1996; Watanabe, 1997); in a third experiment, however, the superiority of glosses on the immediate test had disappeared on the delayed test (Jacobs et al., 1994).
3. Multiple-choice glosses (inferring) led sometimes to better retention than single glosses (giving); see Hulstijn (1989, 1992, exp. 5). Inferring also sometimes led to comparable retention (Hulstijn, exp. 2; Watanabe) and sometimes to lower retention (Hulstijn, exp. 1).
4. Vocabulary explanations provided immediately after the unknown words in the text (giving) led to comparable retention and inferencing (Watanabe).

For intentional learning, Bialystok (1983) and Grace (1998) found that giving was better than inferring, but it should be mentioned that in Bialystok's experiment the difference was possibly caused by incorrect inferencing. Schouten-van Parreren (1980, 1985) found, dependent on the test format, that giving was either better than or comparable to inferring, but she also found that subjects in the meaning-inferred condition had forgotten fewer words than those in the meaning-given condition. Finally, Hulstijn (1989, 1992) found no differences for intentional learning, not even in those cases where the meaning-inferred method led to higher retention under incidental-learning circumstances, a result consistent with Berlyne et al. (1968, exp. 1, delayed test).

RESEARCH OBJECTIVES

As has been shown, there is no conclusive evidence with regard to the supposed superiority of the meaning-inferred method over the meaning-given method. On the other hand, it should be mentioned that the meaning-inferred method has not been realized optimally in all experiments (e.g., incorrect inferencing). Therefore, it is useful to investigate whether an optimized meaning-inferred method leads to better retention than the meaning-given method. This is the first aim of the present experiment. To optimize the meaning-inferred method, three stages were established—namely, inferring, verifying, and memorizing (intentional learning)—and incorrect inferencing was prevented as much as possible by using pregnant sentences.

The second aim of the experiment is to investigate the learning effect of the various stages of the meaning-inferred method compared with each other: inferring, verifying, and memorizing. In this way, the strength of the meaning-inferred method can be determined—a question not yet empirically answered—as well as which perspectives the meaning-inferred method possibly offers for further optimization. The third aim of the experiment is to investigate not only the

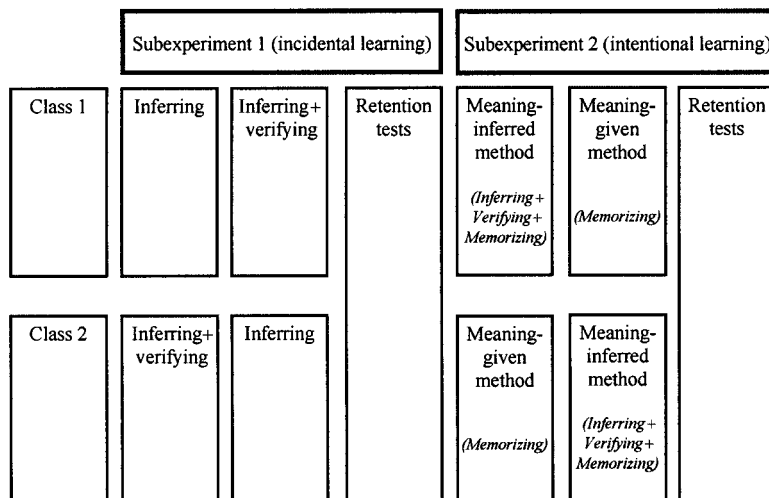


Figure 1. Global experimental design.

learning effect of the meaning-inferred method and the meaning-given method as well as the various stages of the meaning-inferred method but also achievement rate—that is, the learning effect in relation to the amount of time invested. As has been shown, little or no attention has been paid to this efficiency factor (but see Krashen, 1989, appendix 2, which has examined this factor).

METHOD

Design

The global experimental design is shown in Figure 1. Learners of French as a foreign language learned the meanings of four series of French words with the aid of four different learning methods (within-subjects design). The learning methods were (a) inferring, (b) inferring + verifying, (c) inferring + verifying + memorizing (meaning-inferred method), and (d) memorizing (meaning-given method). The experiment took place in the form of two successive subexperiments. The first was an incidental-learning experiment that used learning methods (a) and (b) to prevent the subjects from memorizing. The second was an intentional-learning experiment that used learning methods (c) and (d). For the sake of ecological validity, the experiment took place during the students' regular French lessons, and the target words and the majority of the contexts were taken from existing learning materials. The amount of time invested was left to the students' discretion and was recorded. For each of the two subexperiments, the receptive tests took place 2 weeks later.

Research Questions

The following research questions were addressed.

1. Does learning words through the meaning-inferred method lead to higher retention than learning words through the meaning-given method?
2. What is the learning effect of the various stages of the meaning-inferred method (inferring, verifying, and memorizing) compared to each other?
3. Is there a difference in the amount of time invested between the meaning-inferred method and the meaning-given method?
4. What is the contribution of the various stages of the meaning-inferred method to the total amount of time invested in the meaning-inferred method?
5. What is the achievement rate (i.e., the relation between retention and amount of time invested) of the various conditions and of the various stages of the meaning-inferred method?
6. Is there a relation between correctly or incorrectly inferring and retention? In particular, are correctly inferred words retained better than incorrectly inferred words?

The last research question was included to check the presumed importance of inferring correctly for retention, as an earlier study unexpectedly revealed that contexts that led to improved inferring did not lead to improved retention (Mondria & Wit-de Boer, 1991).

Subjects

Subjects were 38 Dutch students from two third-grade (corresponding to grade 11 in the United States) classes at a school offering general secondary education. The ages of the students were 14–16 years, and all of them had been receiving French lessons 3 hours every week for at least 2.5 years. The same teacher of French taught both classes, and there was no significant difference between them with regard to their knowledge of French as expressed by the students' most recent grades ($p = .058$).

Materials

Target Words and Contexts. Seventy target words (nouns and verbs) with accompanying sentence contexts were selected in the following way. From two French vocabulary textbooks (Daams-Moussault & Blaauw-Holtzappel, 1985, 1986) words were chosen (a) that did not occur in the textbook used in the school concerned (Landgraaf, Brodow, & Durand, 1981), (b) that were not cognates or loan words possibly known by the students, and (c) whose meanings could not be inferred on the basis of words, stems, and affixes already known. Furthermore, the selection did not contain synonymous word pairs or word pairs whose meanings could be inferred from each other.

For each of these words, a pregnant-sentence context (in some cases a definition) was constructed on the basis of contexts found in vocabulary textbooks, learner dictionaries, and monolingual dictionaries. In cases where there

was some doubt as to whether the students would know the words in the context sentence, the translation of these words was given. The degree of pregnancy of the sentence contexts was determined in a pretest conducted at a school that used the same textbook as the school in the experiment. Subsequently, 70 pregnant items (35 nouns and 35 verbs with accompanying contexts) were selected. Items that were pregnant at 100% were not selected because they would have been too easy, which might have had a negative effect on retention in the case of the meaning-inferred method (see Haastrup, 1991; Mondria & Wit-de Boer, 1991). The 70 target words were divided into five sets of 14 words (7 nouns and 7 verbs). Sets 1–4 constituted the words to be learned during the experiment; set 5 was reserved for a control test.

In the learning materials, the target words were underlined, and the translation of context words that might not be known by (some of) the students was given in a smaller font under the words in question. For each set of words, two versions were made that differed with regard to the item order. An example of an item applying to both the meaning-inferred and the meaning-given conditions is given in (1).

- (1) *Notre voiture est assez puissante pour remorquer une caravane.*
“Our car is powerful enough to pull a trailer.”

In both conditions, the Dutch translation of *assez puissante* (*voldoende sterk* “enough powerful”) appeared directly below the words in question in a smaller font, and in the meaning-given condition (but not in the meaning-inferred condition), the Dutch translation of *remorquer* (*trekken* “pull”) was given in the right-hand margin. The complete set of target words appears in the appendix along with English glosses.

Alphabetical Word Lists. For each of the four sets of words to be learned, an alphabetical word list was constructed, containing 14 target words with their meanings and 56 other French words (filler words) with their meanings. The contextually relevant meanings of the target words were taken from French–Dutch dictionaries and supplemented with a few meanings that had been judged as correct on the pretest. At least one filler word was included for each letter of the alphabet. Especially in the immediate alphabetical neighborhood of the target words, filler words were included to generate an intensive search activity and to force the students to closely inspect the word form.

Inferencing Training. To train the students to infer the meanings of unknown words with the aid of the context, a short training session was developed that could be given within a single lesson. The instructions were based on Nation (1990, pp. 162–163), and the exercise sentences were taken from the remainder of the pretest materials.

Test. As a criterion for receptive word knowledge, recall of the meaning (translation) of an L2 word in a new (unhelpful) context was adopted (for a discussion of the concept of successful acquisition, see Henriksen, 1999).

Therefore, for each of the 70 target words a nonpregnant test sentence was constructed, and the nonpregnancy was checked in a pretest at another school. The item order within each test was randomized. An example of a test item is given in (2).

- (2) *Il la remorque.* *remorquer* = _____
 "He pulls it."

Procedures

General Procedure. The experiment consisted essentially of four parts: (a) an inferencing training, (b) an incidental-learning experiment, (c) an intentional-learning experiment, and (d) a test that subsequently measured the knowledge of the type of target words used in the experiment. The procedures for each of the two participating classes were identical except for the order of the two conditions within each of the two subexperiments, which was balanced over the two classes.

Without their knowledge, the students in each class were assigned to one of two subgroups; knowledge of French (most recent grade received) and participation in Latin lessons were controlled. Each of the four subgroups learned a different set of words in each of the four conditions such that each set of words was learned in each condition by one subgroup. This design, depicted graphically in Figure 2, neutralized possibly disturbing effects of condition order, subgroup, and set of words.

The 16 experimental sessions (2 classes \times 8 sessions) took place during the students' regular French lessons. The first session (inferencing training) was led by their French teacher. The other sessions were led by the researcher, assisted during the learning sessions by the French teacher and an assistant who recorded the amount of time spent by the students. This was made possible by asking the students to put their learning materials back into the materials envelope when they had finished.

The experimental materials were distributed at the beginning of each session and were collected at the end of it. The students were not informed of the design of the experiment because of the incidental-learning character of the first subexperiment. For the same reason, the control test was not administered until the end of the experiment. In each condition, students were allotted as much time as they needed for the task, with a maximum of 35 minutes. They were instructed to work accurately and, in the intentional-learning conditions, to continue learning until they really knew the words.

Procedures per Session. The procedures for the eight sessions in class 1 were the following. For class 2, the procedures were identical with the exception of the order of the conditions (see Figure 2).

1. *Inferencing training.* Three weeks before the first subexperiment the students in each class were given an inferencing training using materials developed for this purpose.

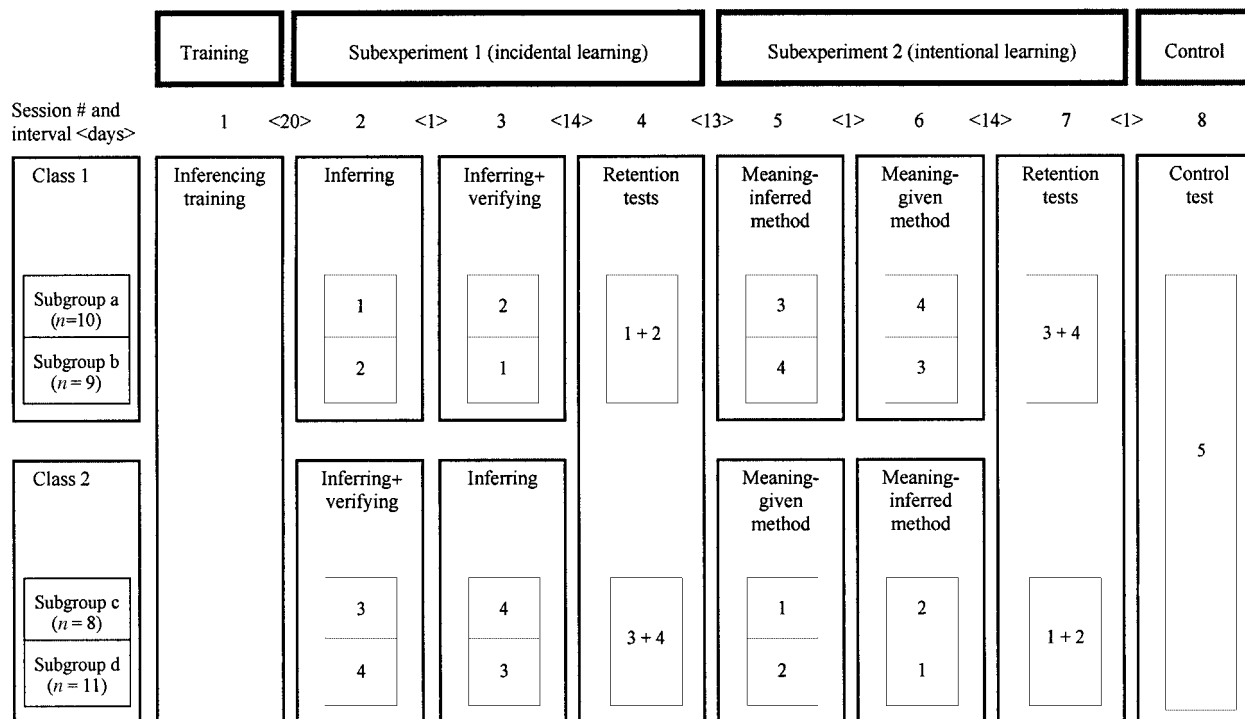


Figure 2. Experimental design (numbers in lighter grey boxes [1–5] refer to sets of target words).

2. *Inferring condition (incidental learning)*: Students were asked to infer and write down the meanings (translations) of the target words.
3. *Inferring + verifying condition (incidental learning)*: The next day students were asked to infer and write down the meanings of the target words and to verify the correctness of the inferred meanings with the aid of the alphabetical word list. If their answer was mentioned in the list, it was considered correct, and they had to indicate this with a tick. In all other cases, the answer was considered incorrect, and they had to correct it using a red pen.
4. *Retention tests (incidental learning)*: The retention tests were administered 2 weeks later. The students were instructed (a) to think back to the sentences that had originally aided them in inferring the meanings of the words, as this could help them; and (b) to write down a meaning (translation) for each word, if necessary by guessing. The time available was 25 minutes. After the tests, they were asked if they had expected to be tested. Only one student responded "yes," but he had no correct answers.
5. *Meaning-inferred method (intentional learning)*: The second subexperiment took place 2 weeks after the first one. The students were asked (a) to infer and write down the meanings of the target words, (b) to verify the correctness of the inferred meanings with the aid of the alphabetical word list and to correct the meaning if necessary with a red pen, and (c) to memorize the inferred and verified meanings of the target words (from French to Dutch). They were also instructed to closely examine the sentences that had originally aided them in inferring the meanings of the words, as this could help them retain and recall the word meanings. At the same time, they were told that they had to know the meanings of the words without any context for the test. To prevent them from talking about the test just before its administration, which might have had a disturbing effect on the measurement, the students were told that the test would be administered after 3 weeks (instead of the actual 2 weeks).
6. *Meaning-given method (intentional learning)*: The day after the meaning-inferred treatment the students were asked to memorize the given meanings of the target words. The instructions for memorizing were comparable to the meaning-inferred method.
7. *Retention tests (intentional learning)*: The retention tests were administered 2 weeks later. The test instructions were comparable to those for the tests for incidental learning, and the time available was also 25 minutes.
8. *Control tests*: The control test was administered the following day. The students were told that the test contained words that they had not come across during the experiment but that they might know. The time available was 10 minutes.

Scoring

Scoring was done by the author and a French teacher. Independently of each other and without any context, they judged the correctness of the answers on the basis of the meanings given in bilingual and monolingual dictionaries. Inflectional forms were treated as lexemes. The mean agreement between judges was .98 (r_s). Differences between judges were resolved by discussion. Correct answers were assigned 1 point and incorrect answers 0 points.

Table 1. Control test and retention scores per condition

Class	<i>n</i>	Control test		Incidental learning				Intentional learning			
				Inferring		Inferring + verifying		Meaning-inferred method		Meaning-given method	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	19	0.11	0.32	0.79	0.85	1.47	1.02	8.47	3.94	6.58	4.26
2	19	0.00	0.00	0.95	0.91	2.84	2.01	4.68	2.98	7.42	3.69
Total	38	0.05	0.23	0.87	0.88	2.16	1.72	6.58	3.94	7.00	3.95

Note. Maximum score = 14.

Data Analysis

The main unit of analysis was a student’s test score on a set of 14 items. Thus, for each student there were four retention scores—one each from (a) the inferring condition, (b) the inferring + verifying condition, (c) the meaning-inferred method, and (d) the meaning-given method. There was also a control-test score for each student. The most important techniques of analysis were multivariate ANOVAs and multiple comparisons based on univariate ANOVAs of difference scores. The within-subjects variable was condition (four). Between-subjects variables were class (two) and subgroup within class (two); the variable class corresponds to the variable condition order. The variable sex had no significant effect in the analyses and will not be discussed for that reason. The effect of the variable grade (for French) will only be reported in those cases in which analyses of covariance provided additional information. The control scores were not used as a covariate because of their low reliability ($\alpha = .06$), which can be accounted for by a floor effect. In addition to these analyses, correlations were calculated between inferring and retention. In all cases, the alpha was set at .05.

RESULTS

General Analysis

The pregnancy of the 56 items, calculated on the basis of the students’ inferring in the four conditions, turned out to be excellent: .89 on average ($SD = 0.11$). The pregnancy was slightly low for only six items (less than .75), the lowest value being .53. The reliability of the retention tests was high: For the sets of words 1–4, the alpha values were .85, .89, .91, and .91, respectively.

The main data of the experiment—that is, the mean retention scores and the mean control-test scores—are given in Table 1. A 4×2 MANCOVA with condition as a within-subjects variable, class as a between-subjects variable,

Table 2. Retention scores per learning session within each subexperiment ($N = 38$)

Subexperiment	Learning session 1		Learning session 2		t^a	df
	M	SD	M	SD		
Incidental learning	1.82	1.84	1.21	0.99	1.72	37
Intentional learning	7.95	3.80	5.63	3.75	5.17*	37

^aPairwise.* $p < .001$ (two-tailed).

and grade as a covariate shows a significant effect of condition, Wilks's $\Lambda = .22$, $F(3, 34) = 41.28$, $p < .001$, and a significant Condition \times Class interaction, Wilks's $\Lambda = .46$, $F(3, 34) = 13.34$, $p < .001$. There was not a significant effect of class, $F(1, 35) = 0.01$, $p = .935$, or grade, $F(1, 35) = 4.00$, $p = .053$. Thus, the different conditions led to differences in retention and, when differences in grade for French are corrected for, there is no significant difference between the two classes.

The interaction between condition and class is clearly visible in the retention figures of the intentional-learning conditions (see Table 1): In class 1, retention scores are higher for the meaning-inferred method than for the meaning-given method, and in class 2, the opposite is the case. The most plausible explanation for this interaction is an effect of condition order: Within a subexperiment, performance in the first condition is relatively better than in the second condition. Evidence for this is the decreasing learning effect within each subexperiment (see Table 2). This decrease was not significant in the first subexperiment ($p = .095$), but it was significant in the second subexperiment ($p < .001$; pairwise, two-tailed t -tests). Possible explanations for this decreasing learning effect are students' decreasing concentration, motivation, or both, caused by their habituation to the (partly identical) tasks.

Research Question 1: Retention in the Meaning-Inferred and Meaning-Given Methods

There was no significant difference in retention between the meaning-inferred method (6.58) and the meaning-given method (7.00) although retention for the meaning-given method was 6% higher (see Table 1). This was shown by a 2×2 MANOVA on the difference scores, with class and subgroup (within class) as between-subjects variables, $F(1, 34) = 0.65$, $p = .427$. There was a significant effect of class, $F(1, 34) = 24.91$, $p < .001$: In class 1, retention was highest for the meaning-inferred method, and in class 2, retention was highest for the meaning-given method. However, this effect can be explained by the effect of condition order, as previously mentioned.

Grade as a covariate had no significant effect on the difference scores.

Thus, it is not the case that one learning method is favorable or unfavorable for students with high or low grades. This is confirmed by the correlations between the students' performance with the meaning-inferred method and their performance with the meaning-given method, which were .80 ($p < .001$) and .63 ($p < .01$) for classes 1 and 2, respectively. In general, students who performed well with the meaning-inferred method also performed well with the meaning-given method and vice versa.

Research Question 2: Retention per Stage of the Meaning-Inferred Method

The differences between the control test (0.05), the inferring condition (0.87), the inferring + verifying condition (2.16), and the meaning-inferred method (6.58) were significant, as shown by three 2×2 MANOVAs (Class \times Subgroup within class) on the respective difference scores: $F(1, 34) = 27.17, p < .001$; $F(1, 34) = 18.62, p < .001$; $F(1, 34) = 71.74, p < .001$. Thus, each stage of the meaning-inferred method contributes to retention: Inferring leads to 6% retention ($[(0.87 - 0.05) / 14]$); verifying, after inferring, leads to 9% retention ($[(2.16 - 0.87) / 14]$); and memorizing, after inferring and verifying, leads to 32% retention ($[(6.58 - 2.16) / 14]$). This means that retention in the meaning-inferred method is for two-thirds of the students (68%) the result of memorizing and for one-third (32%) the result of inferring (13%) and verifying (20%).

The variable class had a significant effect on the difference scores for the meaning-given method and the inferring + verifying condition, $F(1, 34) = 24.37, p < .001$. For the difference scores for the inferring + verifying condition and the inferring condition, only a tendency was found, $F(1, 34) = 3.80, p = .060$. The effect of class can be explained by the effect of condition order.

Research Question 3: Time Investment in the Meaning-Inferred and Meaning-Given Methods

The mean amount of time invested per condition is given in Table 3. (The means are given in seconds for ease of calculation but are not meant to suggest precision.) A 4×2 MANOVA with condition as a within-subjects variable and class as a between-subjects variable showed a significant effect of condition, Wilks's $\Lambda = .08, F(3, 34) = 131.06, p < .001$, such that the time investment per condition differs. There was also a significant effect of class, $F(1, 36) = 12.45, p < .01$, such that class 2 spent more time (16% more) than class 1. Additionally, the interaction Condition \times Class was significant, Wilks's $\Lambda = .29, F(3, 34) = 28.02, p < .001$. An explanation for this is the effect of condition order: The first condition within a subexperiment uses relatively more time. Evidence for this is the significant decrease in time investment from the first to the second learning session within each subexperiment ($p < .001$ and $p < .01$, respectively, in pairwise, two-tailed t -tests); see Table 4. This decrease in time investment—possibly an effect of task habituation, decreasing motivation, or both—goes

Table 3. Time spent per condition (in seconds)

Class	<i>n</i>	Incidental learning				Intentional learning			
		Inferring		Inferring + verifying		Meaning-inferred method		Meaning-given method	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	19	878	183	885	156	1,563	264	941	186
2	19	658	163	956	157	1,716	226	1,639	380
Total	38	768	204	921	158	1,640	254	1,290	461

Note. Maximum = 2,100.

Table 4. Time spent per learning session within each subexperiment ($N = 38$)

Subexperiment	Learning session 1		Learning session 2		<i>t</i> ^a	<i>df</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Incidental learning	917	173	772	195	3.84**	37
Intentional learning	1,601	325	1,329	443	3.71*	37

^aPairwise.

* $p < .01$ (two-tailed). ** $p < .001$ (two-tailed).

hand in hand with a decrease in retention within each subexperiment (see Table 2). However, the decrease in time investment had no disturbing effect on the crucial comparison of the meaning-inferred method and the meaning-given method because the order of these conditions was balanced over the two classes.

Students spent more time (27% more) in the meaning-inferred method than in the meaning-given method, as shown by a 2×2 (Class \times Subgroup within class) MANOVA on the difference scores, $F(1, 34) = 52.75$, $p < .001$. Furthermore, there was a significant effect of class, $F(1, 34) = 32.44$, $p < .001$, such that in class 1 the difference in time investment between the meaning-inferred method and the meaning-given method was bigger than in class 2. This can be explained by the effect of condition order. Grade as a covariate had no effect on the difference score. Thus, the difference in time investment between the two methods had no relationship with students' grade for French.

Research Question 4: Time Investment per Stage of the Meaning-Inferred Method

The differences in time investment between the inferring condition (768 s), the inferring + verifying condition (921 s), and the meaning-inferred method (1,640 s)

were significant, as shown by two 2×2 (Class \times Subgroup within class) MANOVAs on the respective difference scores: $F(1, 34) = 28.87, p < .001$, and $F(1, 34) = 295.81, p < .001$. The addition of a verification stage to the inferring stage leads to an extra time investment, as does the subsequent addition of a memorizing stage. The variable class only had a significant effect on the difference scores for the inferring + verifying condition and the inferring condition, $F(1, 34) = 26.93, p < .001$. This can be explained by the effect of condition order.

The repartition of the time investment in the meaning-inferred method (1,640 s) over the different stages was as follows. Inferring took 481 seconds (this was registered separately), verifying took 153 seconds (inferring + verifying condition [921 s] – inferring condition [768 s]), and memorizing took 1,006 seconds (1,640 – 481 – 153). Thus, the inferring stage takes 29% of the time, the verification stage 9%, and the memorizing stage 61%. This means that the memorizing stage takes up the greater part of the time investment of the meaning-inferred method, an amount of time corresponding to 78% of the time spent on memorizing in the meaning-inferred method.

Research Question 5: Achievement Rate

The relation between retention and time investment per condition has been presented graphically in Figure 3. Most striking is that the meaning-inferred method leads to almost the same retention as the meaning-given method, but the amount of time invested for this retention is considerably larger (27% more time). Additionally, this figure shows that each stage of the meaning-inferred method (inferring, verifying, and memorizing) leads to retention and that this is connected with a distinct time investment.

To compare the efficiency of the conditions and the stages of the meaning-inferred method, achievement rates were calculated, expressed as the average number of words acquired per minute (see Krashen, 1989). Table 5 shows that the achievement rate of the meaning-given method (.32) is higher than that of the meaning-inferred method (.24). To learn a word, the meaning-given method uses 26% less time than the meaning-inferred method. Within the meaning-inferred method, the achievement rate of intentional learning (memorizing = .37) is clearly higher than that of incidental learning (inferring + verifying = .14). In incidental learning, verifying (.51) has a much higher achievement rate than inferring (.06).

Research Question 6: Relation between Correctly or Incorrectly Inferring and Retention

The relation between correctly or incorrectly inferring the meaning of a word and its retention was investigated by calculating the retention of correctly and incorrectly inferred words separately. In the inferring condition, the retention of correctly inferred words was, as expected, higher than the (correct) retention of incorrectly inferred words (7% vs. 2%; one incorrectly inferred word

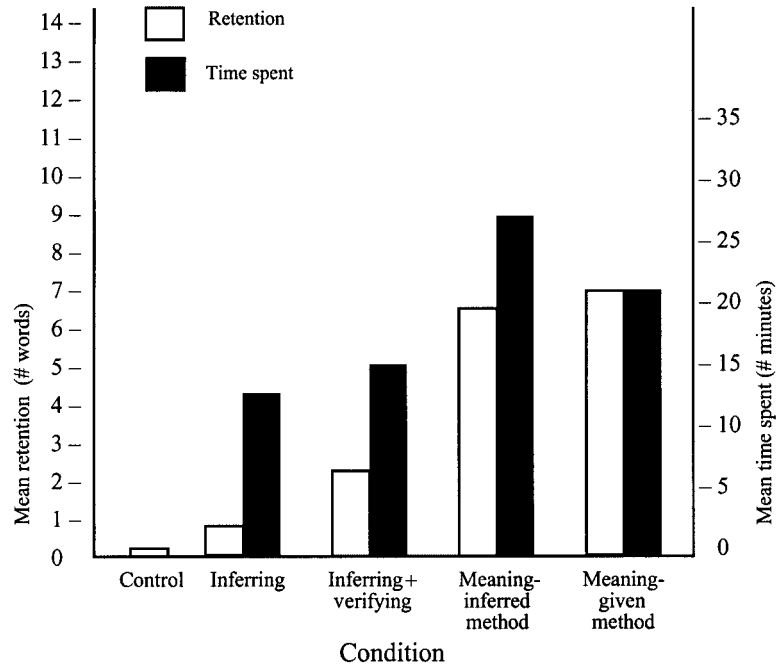


Figure 3. Retention and time spent, per condition.

Table 5. Achievement rate per condition and per stage of the meaning-inferred method ($N = 38$)

Condition/stage	Words/minute
Condition	
Inferring condition	.06 ^a
Inferring + verifying condition	.14 ^a
Meaning-inferred method	.24 ^a
Meaning-given method	.32 ^a
Stage of the meaning-inferred method	
Inferring	.06 ^a
Verifying	.51 ^b
Memorizing	.37 ^c

^aCalculation: (mean retention score - mean control score)/mean time spent.

^bCalculation: (mean retention score for inferring and verifying - mean retention score for inferring)/(mean time spent on inferring and verifying - mean time spent on inferring).

^cCalculation: (mean retention score for meaning-inferred method - mean retention score for inferring and verifying)/(mean time spent on meaning-inferred method - mean time spent on inferring and verifying).

was unexpectedly translated correctly on the retention test). There was no correlation between inferring and retention.

In the inferring + verifying condition, the (correct) retention of incorrectly inferred words was comparable to that of correctly inferred words (16% vs. 15%). This is a clear indication that the correction of incorrectly inferred meanings brings about a greater learning effect than the confirmation of the correctness of inferred meanings. There was no correlation between inferring and retention.

In the meaning-inferred method, correctly inferred words were considerably better retained than incorrectly inferred words (50% vs. 31%). There was a weak correlation between inferring and retention: r_s (incorrect = 0; correct = 1) = .13, $p < .01$. Additionally, the percentage of 50 makes it clear that even if retention in the meaning-inferred method is calculated on the basis of the correctly inferred words, retention is still not higher than in the meaning-given condition (50%).

DISCUSSION AND CONCLUSIONS

Meaning-Inferred Method versus Meaning-Given Method

Learning words through the meaning-inferred method leads to a similar level of retention as learning words through the meaning-given method (47% and 50% retention, respectively). This is contrary to expectations. It had been expected that the meaning-inferred method would lead to higher retention than the meaning-given method. This expectation was based on the assumed deeper, elaborative processing as a result of inferring and verifying, thus creating all kinds of links between the word to be learned, its meaning, the context, and the knowledge of the learner. Additionally, the expectation was consistent with the involvement load hypothesis (Laufer & Hulstijn, 2001) because of its search and evaluation activities.

This unexpected result cannot be accounted for by the way in which the meaning-inferred method has been operationalized. That is, the contexts were so pregnant that inferring wrongly was prevented to a large extent, and even after elimination of the incorrectly inferred words, the meaning-inferred method did not lead to higher retention than the meaning-given method. Furthermore, the form of the verification stage was such that it generated genuine, cognitive activity without the risk of making mistakes.

The explanation that the students lacked familiarity with the meaning-inferred method also cannot account for the results. The fact is that, before the crucial comparison between the two methods, students were familiarized with the specific stages of the meaning-inferred method by means of an inferencing training and two conditions in which they (a) inferred and (b) inferred and verified. Moreover, the amount of time to be spent was left up to them so that a possible unfamiliarity with the meaning-inferred method would not be at the expense of the amount of time to be spent usefully. The conclusion is also robust

in other respects. Much attention had been paid to the validity of the experiment, and the conclusion holds both for students with a good knowledge of French and for those with less knowledge of French.

In view of the preceding discussion, the explanation for the fact that the meaning-inferred method is not superior probably must be sought in the method itself. In particular, it must be explained why the meaning-inferred method does not lead to improved retention in spite of its two extra stages (inferring and verifying)—stages that entail search and evaluation activities and that both lead to retention. The following two explanations are possible.

A first explanation is that the quantity of the memorizing activity in the meaning-inferred method is less than that in the meaning-given method, as has been found on the basis of the time recordings. There are at least three possible accounts for this. The first is that inferring and verifying already lead to a certain level of retention, and as a result students need to invest less time in memorizing to achieve the intended level of knowledge. The second possibility is that inferring and verifying together already take so much time that students are less motivated to invest the same amount of time in memorizing as in the meaning-given method. The third possibility is that inferring a word correctly may give students the impression that they know the word already (Mondria & Wit-de Boer, 1991), and as a result they will spend less time memorizing.

A second explanation for the fact that retention in the meaning-inferred method is not higher in spite of the two extra stages is that the quality of the memorizing activity in the meaning-inferred method is inferior to that in the meaning-given method. The fact that inferencing draws attention mainly to the association of context and meaning might influence subsequent memorization in such a way that less attention is paid to the association of word (form) and meaning that is essential for retention.

The result found corresponds to the delayed-test results in Berlyne et al. (1968, exp. 1), to the isolated-words test in Schouten-van Parreren (1980, 1985), and to Hulstijn (1989, 1992), who, in an intentional-learning situation, also did not find a significant difference between the two methods. However, the result conflicts with Schouten-van Parreren's results for the nonpregnant-sentences test, with Bialystok (1983, exp. 2), and with Grace (1998), where the meaning-inferred method was inferior to the meaning-given method, possibly as a consequence of inferring incorrectly. A comparison with the result found by Schouten-van Parreren that learners on the delayed test had forgotten fewer words in the meaning-inferred method than in the meaning-given method is not possible because in the present experiment no immediate test was administered.

Viewed in a wider perspective, the experiment makes clear that actions that are supposed to lead to more (relevant) cognitive activity—"elaboration" (Anderson, 1990) or "search" and "evaluation" (Laufer & Hulstijn, 2001)—do not automatically lead to better retention. This conclusion is consistent with Eysenck (1982), who concluded that "the typical finding is that intentional

learners not performing an orienting task recall at least as well as intentional learners who are given a semantic orienting task” (p. 203).

With regard to the amount of time spent, it can be concluded that the meaning-inferred method takes considerably more time (about 25% more) than the meaning-given method, which conforms to the expectations. As retention in both methods is similar, the achievement rate of the meaning-given method is higher than that of the other method: Per word learned, the meaning-given method uses about 25% less time.

Meaning-Inferred Method

Each of the various stages of the meaning-inferred method leads to retention: Inferring leads to 6% retention, verifying leads to 9% retention, and memorizing leads to 32% retention. With regard to inferencing, the experiment makes clear that its retention effect (6%) should not be overestimated: The correct inferencing of the meaning of a word is by no means a guarantee of retention (Jenkins & Dixon, 1983). Moreover, this percentage was reached under favorable circumstances: pregnant contexts in combination with conscious inferencing and writing down the inferred meaning. On the other hand, the learning effect of inferencing is greater than the retention as measured by the open, nonpregnant-sentences test. According to the multiple-choice tests, which were administered as an additional check after each nonpregnant-sentences test but which were not reported in this article to not complicate the description, more is retained (20%) even though this is at a relatively low level of knowledge.

An objective comparison of the incidental-learning effect of inferencing in the present experiment with that in other studies is not possible because this experiment differs in one or more crucial factors. Examples include: (a) the type of activity (explicit inferencing and noting vs. inferencing for the purpose of text comprehension; inferencing with the aid of context vs. inferencing with the aid of context and multiple-choice glosses; reading vs. reading while the text is read aloud; and presence of a pretest that possibly draws attention to vocabulary), (b) the type of material (isolated sentences vs. a coherent text; number, type, and frequency of the target words), (c) the type of learner (secondary school vs. university), and (d) the type of test (open vs. multiple-choice; retention interval; repeated tests, as a result of which the test actually becomes part of the experimental treatment). It is clear, however, that the receptive retention percentages mentioned (6% for the open test and 20% for the multiple-choice test) are within the range indicated by other L2 experiments. Horst, Cobb, and Meara (1998), for example, concluded that retention in seven previous studies (six of which contained a multiple-choice test) was 8% on average, whereas they found in their own experiment 20% retention for the multiple-choice test. The difference with the relatively high retention percentages found by Rott (1999)—22% and 14% on the open test after 1 and 4 weeks, respectively—may be mainly accounted for by the fact that Rott used

two different contexts per word, separated by an interval of 1 week and followed by two immediate tests (open and multiple-choice). Moreover, the retention percentages found for inferencing are consistent with the results of L1 research into incidental learning on the basis of reading. Nagy (1997), for example, found in four experiments (three of which had an open test) retention percentages between 5% and 14%, and Swanborn and De Glopper (1999) found a retention percentage of 15% on the basis of a meta-analysis of 20 experiments (14 of which had an open test).

The retention effect of verification (9%) is comparable to that of inferencing (6%). Additionally, the learning effect of the correction of an incorrectly inferred meaning is greater than the learning effect of the confirmation of a correctly inferred meaning (cf. Mondria & Wit-de Boer, 1991). As a result, the (correct) retention of incorrectly inferred words is, after verification, similar to that of correctly inferred words after verification (16% and 15%, respectively). A comparison with the related experiments by Luppescu and Day (1993) and Knight (1994) on inferencing and looking up is not possible because it cannot be ruled out that, in these experiments, retention as a result of looking up must be interpreted as the result of inferencing and looking up.

Finally, with 32% retention, memorization accounts for the major part (two-thirds) of retention in the meaning-inferred method. Therefore, there is no reason for relatively high expectations with regard to the retention effect of inferencing and verification. Furthermore, the learning effect of memorization is greater for correctly inferred words than for incorrectly inferred ones. Although retention for both types of words was comparable after inferring and verifying, there is a considerable difference in retention—50% versus 31%, respectively—after the addition of a memorization stage. The positive learning effect of inferring correctly can thus be found back in the retention of the meaning-inferred method as a whole. However, the correlation between inferencing and retention is weak ($r_0 = .13$) as a consequence of (a) not retaining correctly inferred words and (b) the relatively high correct retention of incorrectly inferred words (cf. Mondria & Wit-de Boer, 1991).

With regard to the time spent on the meaning-inferred method, the memorization stage accounts for the greater part of it (61%). The inferencing stage takes 29% of the time, and the verification stage takes 9%. The combination of these data with the retention data leads to the conclusion that, within the meaning-inferred method, the achievement rate of intentional learning (memorizing) is clearly higher than that of incidental learning (inferring + verifying). Furthermore, within the incidental-learning part, the achievement rate of verifying appeared to be much higher than that of inferring, a result that has not been shown by earlier research.

Implications for Foreign Language Teaching

When the efficiency of the vocabulary acquisition process is the main objective, the meaning-given method is preferable to the meaning-inferred method

because less time has to be invested for a similar level of retention. Additional advantages of the meaning-given method are that the method sets less stringent requirements for the learning materials (contexts) and no requirements for the inferencing skills of the students (for an overview of the limitations of inferring from context, see Huckin & Coady, 1999).

With regard to incidental vocabulary acquisition through reading, many studies have shown that retention as a result of exclusively inferring is rather low. If one wants to improve this, it is advisable to add a verification stage, which takes little time and improves retention considerably as well as the achievement rate. At the same time, the retention of incorrectly inferred meanings is counteracted. However, for substantial retention it is desirable to add a memorization stage after inferring and verifying, which improves the achievement rate once again (see Hulstijn, 2001, and Hulstijn et al., 1996, who also argued for intentional learning as a follow-up to incidental learning, and Mondria & Mondria-de Vries, 1994, for an efficient and practical form for this intentional-learning stage).

Suggestions for Further Research

Further research could investigate whether the conclusions with regard to these two methods also hold for the learning of other types of words—for example, words with morphological clues and parts of speech other than nouns and verbs. Moreover, the validity of the conclusions could be investigated for students at lower levels, who have more difficulty with inferencing (Knight, 1994) as well as with memorization.

(Received 25 September 2002)

REFERENCES

- Anderson, J. R. (1990). *Cognitive psychology and its implications* (3rd ed.). New York: Freeman.
- Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Reading*, 7, 15–32.
- Berlyne, D. E., Carey, S. T., Lazare, S. A., Parlow, J., & Tiberius, R. (1968). Effects of prior guessing on intentional and incidental paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 7, 750–759.
- Bialystok, E. (1983). Inferencing: Testing the “hypothesis-testing” hypothesis. In H. W. Seliger & M. H. Long (Eds.), *Classroom oriented research in second language acquisition* (pp. 104–123). Rowley, MA: Newbury House.
- Daams-Moussault, A., & Blaauw-Holtzappel, F. M. M. (1985). *Nouveau vocabulaire français* (4th ed.) [New French vocabulary]. Zeist, the Netherlands: NIB.
- Daams-Moussault, A., & Blaauw-Holtzappel, F. M. M. (1986). *Nouveau vocabulaire de base PLUS* [New basic French vocabulary PLUS]. Zeist, the Netherlands: NIB.
- Ellis, N. C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning*, 8, 103–128.
- Eysenck, M. W. (1982). Incidental learning and orienting tasks. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 197–228). San Diego, CA: Academic Press.
- Fischer, U. (1994). Learning words from context and dictionaries: An experimental comparison. *Applied Psycholinguistics*, 15, 551–574.
- Forlano, G., & Hoffman, M. N. H. (1937). Guessing and telling methods in learning words of a foreign language. *Journal of Educational Psychology*, 28, 632–636.

- Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, 21, 225–241.
- Gass, S. (1999). Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition*, 21, 319–333.
- Gershman, S. J. (1970). *Foreign language vocabulary learning under seven conditions*. Unpublished doctoral dissertation, Columbia University, New York.
- Grace, C. A. (1998). Retention of word meanings inferred from context and sentence-level translations: Implications for the design of beginning-level CALL software. *The Modern Language Journal*, 82, 533–544.
- Haastруп, K. (1991). *Lexical inferencing procedures or talking about words: Receptive procedures in foreign language learning with special reference to English*. Tübingen: Gunter Narr.
- Haynes, M. (1993). Patterns and perils of guessing in second language reading. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 46–62). Westport, CT: Ablex.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21, 303–317.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond *A Clockwork Orange*: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11, 207–223.
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition*, 21, 181–193.
- Hulstijn, J. H. (1989). Het afleiden van de betekenis van vreemdtaalige woorden uit de context: Experimenteel onderzoek naar het effect van vier vormen van steun [Learning word meanings from context: An experimental comparison of the effect of four orienting cues]. *Toegepaste Taalwetenschap in Artikelen*, 34, 13–25.
- 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. (2001). Intentional and incidental vocabulary learning: A reappraisal of elaboration, rehearsal, and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). New York: Cambridge University Press.
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80, 327–339.
- 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.
- Jenkins, J. R., & Dixon, R. (1983). Vocabulary learning. *Contemporary Educational Psychology*, 8, 237–260.
- Jenkins, J. R., Matlock, B., & Slocum, T. A. (1989). Two approaches to vocabulary instruction: The teaching of individual word meanings and practice in deriving word meaning from context. *Reading Research Quarterly*, 24, 215–235.
- Jenkins, J. R., Pany, D., & Schreck, J. (1978). *Vocabulary and reading comprehension: Instructional effects* (Tech. Rep. No. 100). Champaign: University of Illinois, Center for the Study of Reading.
- Johnson, D. M., & Stratton, R. P. (1966). Evaluation of five methods of teaching concepts. *Journal of Educational Psychology*, 57, 48–53.
- Kelly, P. (1990). Guessing: No substitute for systematic learning of lexis. *System*, 18, 199–207.
- Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal*, 78, 285–299.
- Koster, L. (1989). Woorden leren in een vreemde taal: De relatie tussen instructie en retentie [Acquiring a foreign language vocabulary: The relation between instruction and retention]. *Toegepaste Taalwetenschap in Artikelen*, 35, 74–82.
- Krantz, G. (1991). *Learning vocabulary in a foreign language: A study of reading strategies*. Göteborg, Sweden: Acta Universitatis Gothoburgensis.
- Krashen, S. D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73, 440–464.
- Landgraaf, W., Brodow, U., & Durand, T. (1981). *On y va tous!* [We go for it!]. Amsterdam: Meulenhoff Educatief.
- Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language*

- vocabulary acquisition: A rationale for pedagogy* (pp. 20–34). New York: Cambridge University Press.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26.
- Laufer, B., & Sim, D. D. (1985). Taking the easy way out: Non-use and misuse of clues in EFL reading. *English Teaching Forum*, 23(2), 7–10, 20.
- Levine, A., & Reves, T. (1990). Does the method of vocabulary presentation make a difference? *TESL Canada Journal*, 8, 37–51.
- Luppescu, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning*, 43, 263–287.
- McDaniel, M. A., & Pressley, M. (1984). Putting the keyword method into context. *Journal of Educational Psychology*, 76, 598–609.
- Miller, G. A. (1991). *The science of words*. New York: Scientific American Library.
- Mondria, J.-A. (1996). *Vocabulaireverwerving in het vreemde-talenonderwijs: De effecten van context en raden op de retentie*. [Vocabulary acquisition in foreign language teaching: The effects of context and guessing on retention]. Unpublished doctoral dissertation, University of Groningen, the Netherlands.
- Mondria, J.-A., & Mondria-de Vries, S. (1994). Efficiently memorizing words with the help of word cards and “hand computer”: Theory and applications. *System*, 22, 47–57.
- 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. (1997). On the role of context in first- and second-language vocabulary learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 64–83). New York: Cambridge University Press.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Rowley, MA: Newbury House.
- Qian, D. D. (1996). ESL vocabulary acquisition: Contextualization and decontextualization. *The Canadian Modern Language Review*, 53, 120–142.
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners’ incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, 21, 589–619.
- Schatz, E. K., & Baldwin, R. S. (1986). Context clues are unreliable predictors of word meanings. *Reading Research Quarterly*, 21, 439–453.
- Schouten-van Parreren, M. C. (1980). Vreemdtalige woorden: De betekenis onmiddellijk geven of eerst laten raden? [Foreign language vocabulary teaching: Getting or guessing translations of new words?]. *Toegepaste Taalwetenschap in Artikelen*, 7, 156–173.
- Schouten-van Parreren, M. C. (1985). *Woorden leren in het vreemde-talenonderwijs* [Learning words in foreign language teaching]. Apeldoorn, the Netherlands: Van Walraven.
- Stein, M. J. (1993). The healthy inadequacy of contextual definition. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 203–212). Westport, CT: Ablex.
- 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). Mahwah, NJ: Erlbaum.
- Swanborn, M. S. L., & De Glosper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69, 261–285.
- Thomas, M. H., & Dieter, J. N. (1987). The positive effects of writing practice on integration of foreign words in memory. *Journal of Educational Psychology*, 79, 249–253.
- Van Parreren, C. F. (1967). Psychologische factoren bij het verwerven van de woordenschat van een vreemde taal [Psychological factors in foreign language vocabulary acquisition]. *Levende Talen*, 239, 159–169.
- Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 19, 287–307.

APPENDIX

Table A1. Target words and glosses

Set	French target word	Dutch gloss	English gloss
Set 1			
Nouns	<i>le chandail</i>	<i>de trui</i>	the sweater
	<i>la cicatrice</i>	<i>het litteken</i>	the scar
	<i>le cuivre</i>	<i>het koper</i>	brass
	<i>les dégâts</i>	<i>de schade</i>	the damage
	<i>l'égout</i>	<i>de riolering</i>	the sewerage
	<i>le ruban</i>	<i>het lint</i>	the ribbon
	<i>le ver</i>	<i>de worm</i>	the worm
Verbs	<i>bousculer</i>	<i>omverlopen</i>	to knock over
	<i>éblouir</i>	<i>verblinden</i>	to blind
	<i>ébranler</i>	<i>doen trillen</i>	to shake
	<i>hurler</i>	<i>huilen</i>	to howl
	<i>licencier</i>	<i>ontslaan</i>	to dismiss
	<i>remorquer</i>	<i>trekken</i>	to pull
	<i>retentir</i>	<i>naklinken</i>	to resound
Set 2			
Nouns	<i>l'averse</i>	<i>de regenbui</i>	the downpour
	<i>la corbeille</i>	<i>de mand</i>	the basket
	<i>l'écurie</i>	<i>de (paarden)stal</i>	the stable
	<i>le gouffre</i>	<i>de afgrond</i>	the abyss
	<i>la griffe</i>	<i>de klauw</i>	the claw
	<i>la paupière</i>	<i>het ooglid</i>	the eyelid
Verbs	<i>le sigle</i>	<i>de afkorting</i>	the abbreviation
	<i>décéler</i>	<i>ontdekken</i>	to discover
	<i>s'écrouler</i>	<i>instorten</i>	to collapse
	<i>étinceler</i>	<i>schitteren</i>	to glitter
	<i>gratter</i>	<i>krabben</i>	to scratch
	<i>ramper</i>	<i>kruipen</i>	to crawl
	<i>rôder</i>	<i>zwerven</i>	to roam
<i>susciter</i>	<i>opwekken</i>	to evoke	
Set 3			
Nouns	<i>la charrue</i>	<i>de ploeg</i>	the plow
	<i>la cuisse</i>	<i>de dij</i>	the thigh
	<i>l'enfer</i>	<i>de hel</i>	the hell
	<i>le gibier</i>	<i>het wild</i>	the game
	<i>l'haleine</i>	<i>de adem</i>	the breath
	<i>le noyau</i>	<i>de pit</i>	the stone
Verbs	<i>la pellicule</i>	<i>het filmpje</i>	the film
	<i>foncer sur</i>	<i>zich storten op</i>	to charge at
	<i>gaspiller</i>	<i>verspillen</i>	to waste
	<i>ménager</i>	<i>regelen</i>	to arrange
	<i>se noyer</i>	<i>verdrinken</i>	to drown
	<i>ronger</i>	<i>knagen</i>	to gnaw
	<i>taquiner</i>	<i>plagen</i>	to tease
<i>virer</i>	<i>gireren</i>	to transfer	

Table A1. Continued

Set	French target word	Dutch gloss	English gloss
Set 4			
Nouns	<i>la grenouille</i>	<i>de kikker</i>	the frogg
	<i>le mazout</i>	<i>de stookolie</i>	the fuel oil
	<i>les noces</i>	<i>de bruiloft</i>	the wedding
	<i>la poubelle</i>	<i>de vuilnisbak</i>	the garbage can
	<i>le serment</i>	<i>de eed</i>	the oath
	<i>le torchon</i>	<i>de theedoek</i>	the dishtowel
	<i>le verger</i>	<i>de boomgaard</i>	the orchard
Verbs	<i>bâiller</i>	<i>gapen</i>	to yawn
	<i>échouer</i>	<i>zakken</i>	to fail
	<i>s'égarer</i>	<i>verdwalen</i>	to get lost
	<i>lécher</i>	<i>likken</i>	to lick
	<i>ramer</i>	<i>roeien</i>	to row
	<i>refouler</i>	<i>bedwingen</i>	to restrain
	<i>sévir</i>	<i>heersen</i>	to be rampant
Set 5			
Nouns	<i>le bourreau</i>	<i>de beul</i>	the executioner
	<i>le comble</i>	<i>het toppunt</i>	the peak
	<i>les déchets</i>	<i>het afval</i>	the garbage
	<i>l'évier</i>	<i>de gootsteen</i>	the kitchen sink
	<i>le foin</i>	<i>het hooi</i>	hay
	<i>le poêle</i>	<i>de kachel</i>	the stove
	<i>le versant</i>	<i>de helling</i>	the hillside
Verbs	<i>agacer</i>	<i>irriteren</i>	to irritate
	<i>débarrasser</i>	<i>ontruimen</i>	to clear
	<i>se déclencher</i>	<i>in werking treden</i>	to break out
	<i>dépouiller</i>	<i>beroven</i>	to rob
	<i>épuiser</i>	<i>uitputten</i>	to exhaust
	<i>jaillir</i>	<i>spuiten</i>	to spout
	<i>tremper</i>	<i>natmaken</i>	to wet