Retention of inferred and given word meanings: experiments in incidental learning

Hulstijn, J.H.

Published in:
Vocabulary and applied linguistics

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
1. INTRODUCTION

One of the much debated issues in foreign/second language (henceforth L2) pedagogy presently is the following question: What is more effective and efficient for L2 vocabulary acquisition: (a) letting L2 learners infer the meaning of an unknown word occurring in an L2 text, using the information contained in the context, or (b) simply giving learners the meaning of the unknown word right away, e.g. by providing a translation in the mother tongue (L1) or a synonym in L2? Some researchers (e.g. Bialystok, 1983; Nation, 1982; Nation and Coady, 1988; Schouten-van Parreren, 1985, 1986) have suggested that new words can best be learnt when presented in texts and when their meaning must be inferred from context by learners themselves. This point of view is based on two assumptions (Craik and Tulving, 1975; Jacoby, 1978; Jacoby and Craik, 1979; Jacoby and Craik, 1979; Jacoby, Craik and Begg, 1979):

(a) When subjects have to infer or induce the solution of a problem, they will invest more mental effort than when they are given the solution to the problem.

(b) Information that has been attained with more mental effort can later be better retrieved and recalled than information that has been attained with less mental effort.

On the basis of these assumptions, one could construct a 'mental effort' hypothesis, which predicts that the retention of an inferred word meaning will be higher than the retention of a given word meaning.

Other researchers, however, object to the inferring-from-context method, mainly on practical, pedagogical grounds (Bensoussan...
Vocabulary and Applied Linguistics

and Laufer, 1984; Carnine, Kameenui and Coyle, 1984; Laufer and Sim, 1985; McKeown, 1985; Kelly, 1989; Koster, 1985; Stip and Hulstijn, 1986). They argue that:

(a) context seldom offers enough information for the inferring method to be successfully applied;
(b) the inferring method has the inherent risk that learners will make wrong inferences, and hence learn the wrong meaning (which they have to 'unlearn' subsequently);
(c) the inferring method only works well with learners who have good problem-solving skills.

2. EXPERIMENTS

2.1. Research objectives

The present study was designed to make an empirical contribution to the discussion of the advantages and disadvantages of the inferring method. The study had two objectives. The first objective was to provide empirical support for the prediction, derived from the mental effort hypothesis, that the retention of inferred meanings is higher than the retention of given word meanings. The second objective was to investigate whether the potential disadvantages of the inferring method could be reduced by means of a multiple-choice procedure (MC procedure), directing and channelling the inferring process. This procedure presents the L2 reader/learner with the question: 'What is the meaning of word X in line Y? Pay attention to the context and choose from the following alternatives.' Then, several descriptions of the meaning of word X are given (e.g. translations in L1 or synonyms in L2), one being correct, the others being incorrect. An essential requirement is that the incorrect descriptions (the distractors) should not be too plausible nor too implausible. When the distractors are too plausible, learners might infer the wrong meaning. When they are too implausible, learners might not take them seriously, in other words, learners would not be forced to make a serious effort to process the relevant information deep enough. Hence, the correct meaning would after all not be inferred, as intended, but given. In summary, the idea behind the MC procedure was to retain the cognitive advantages of the inferring method while reducing the three risks mentioned above, on the basis of the following reasoning:

(a) since the MC procedure gives learners/readers a cue, it offers a compensation for the limited information provided by the context;
(b) with the MC procedure, the chance that learners will infer wrong meanings is reduced to the chance of selecting a/the distractor;
(c) since the MC procedure gives learners/readers a cue, it may also be appropriate for learners/readers with less well-developed problem-solving skills.

2.2. Design

Five experiments were conducted. In Experiments I and II, three meaning-to-be-inferred procedures (Multiple Choice, Concise Context, and No-Cue/Control) were compared with one meaning-given procedure (Translation/Synonym), in order to investigate the first research question, whether inferred meanings are better recalled than given meanings. They were also compared with each other to investigate the second research question, pertaining to the relative merits of the MC procedure. In Experiments III, IV and V, only two procedures were compared, namely Multiple Choice (meaning to be inferred) and Synonym (meaning given).

The common features in design and method between these five experiments are presented in this section. In subsequent sections, the experiments will be described individually. In all the experiments, subjects were given a reading comprehension task, followed by one or more unexpected post-tests. Subjects were instructed to read a text and to answer six multiple-choice comprehension questions. The text was a four-page (907 words) Dutch text about the function of advertising for producers and consumers. All the words in the text, except the target words, were high frequency words (belonging to the 2,000 most frequent words in Dutch). The twelve target words were low frequency words (in Experiments I, III and V, with non-native subjects) or pseudo-words (in Experiments II and IV, with native subjects). The meaning of these words was sometimes given in the margin of the text (Translation/Synonym condition); sometimes cues were
given about these twelve words, so that Ss could infer their meaning (Multiple-Choice and Concise Context conditions); sometimes no information was given at all (No-Cue control condition). Since the task was presented as a reading comprehension task and subjects did not know they would be later tested on their knowledge of the twelve target words, the experiments can be said to have created an incidental learning situation. That is, with the unexpected post-tests we wanted to assess the 'net' retention effect of the information that had been processed during the text comprehension task. If we had told Ss in advance that they would be tested afterwards, they would have invoked all kinds of rehearsal and memorisation techniques, interfering with the effect of the initial processing of the word meanings. Thus, the incidental learning design was adopted because we wanted to assess possible differences in retention between experimental conditions, knowing in advance from the evidence in the literature that, in incidental learning tasks in which each target is presented only once, retention height is very low indeed (Eysenck, 1982).

Experiments I and II had an independent-groups design. Each subject group served in one condition only. For Experiments III, IV and V, however, a within-subject design was adopted. This was done in order to control for potential individual differences (e.g. differences in L2 knowledge, especially vocabulary knowledge) which might interfere with treatment effects.

In Experiment I, a 75-item vocabulary pre-test was administered. These 75 items included the twelve target words used in the experimental task. The mean performance on these 12 words was 1.7. No significant differences existed between the four treatment groups in this experiment.

In Experiment II, pseudo-words were used as targets and native speakers as subjects, so that we could safely assume that no subject knew any of the target words in advance, whereas all subjects knew all non-target words in the text.

In Experiments III, IV and V, the problem of pre-knowledge was dealt with in the following way. Half of the subjects were given the even targets in the Synonym condition and the odd targets in the Multiple-Choice condition; the other half of the subjects were given the even targets in the MC condition and the odd targets in the Synonym condition. Moreover, Experiment IV used pseudo-words as targets and native speakers as subjects (see Hulstijn (1989) for a justification of this research method, using parallel experiments with natural targets and real L2 learners on the one hand, and semi-artificial materials and native speakers on the other hand).

2.3. Experiment I

This experiment was conducted with 65 adult learners of Dutch as a second language. All subjects in the Translation group had Turkish as their L1. There were four groups:

(a) **Translation group.** For each target word, a Turkish translation was given in the margin of the text.

(b) **Concise Context group.** For each target word, a sample sentence was given in the margin, providing a concise context, from which the target's meaning could easily be inferred.

(c) **Multiple-Choice group.** For each target, four alternatives were given in the margin, with the instruction to choose the correct alternative by using the information in the text.

(d) **No-Cue/Control group.** This group was given no information in the margin about the targets.

After having read the text and completed the comprehension questions, subjects were given two unexpected post-tests. In the Meaning test, they had to write down the meanings of the twelve target words. In the Form test, Ss had to fill in the target words in the original text (in open spaces where they had originally occurred).

**Results**

These are given in Table 11.1. The mean retention figures, as expected with an incidental learning task, are very low indeed. Important, however, is the finding that, on the Meaning test, the differences between conditions ran counter to our two predictions: (a) when the meaning of the targets was given (Translation), retention was better than when the meaning had to be inferred, and (b) the MC procedure did not yield a higher retention than the other two inferring conditions (Concise Context and No-Cue/Control).

2.4. Experiment II

Subjects in this experiment were 98 adult Dutch native speakers. There was one meaning-given condition (Synonym) and three
Table 11.1 Experiment I: mean retention of targets (max. = 12)

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Meaning test</th>
<th>Form test**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>17</td>
<td>3.1*</td>
<td>1.4</td>
</tr>
<tr>
<td>Concise Context</td>
<td>16</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>17</td>
<td>2.0</td>
<td>1.5</td>
</tr>
<tr>
<td>No Cue/Control</td>
<td>15</td>
<td>0.9</td>
<td>1.9</td>
</tr>
</tbody>
</table>

* Translation group significantly higher than the other groups.
** No significant group differences.

Table 11.2 Experiment II: mean retention of targets (max. = 12)

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Meaning test</th>
<th>Form test***</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synonym</td>
<td>24</td>
<td>2.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Concise Context</td>
<td>25</td>
<td>3.7</td>
<td>2.0</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>24</td>
<td>2.5</td>
<td>1.8</td>
</tr>
<tr>
<td>No Cue/Control</td>
<td>25</td>
<td>2.2</td>
<td>3.8</td>
</tr>
</tbody>
</table>

* No significant differences between groups.
** No Cue/Control significantly higher than all other conditions.
*** No significant differences between groups.

Results

These are given in Table 11.2. No significant differences between groups were found on either post-test. However, Ss in the No Cue/Control group gave more erroneous responses than Ss in the other groups. They had often inferred possible, albeit incorrect, meanings. This result empirically illustrates the second objection against the inferring method mentioned in the introduction, namely that it has the inherent risk that learners will make wrong inferences, and hence learn wrong meanings.

2.5. Experiment III

This experiment was conducted with 45 adult Dutch L2 learners. There were two cue conditions, Synonym and MC. Six target words (the even or the odd targets) were explained, in the margin, with a Dutch synonym; for the remaining six target words (the odd or even ones respectively), four alternative Dutch 'synonyms' were given, one being the correct synonym, the other three being distractors. There was one (unexpected) post-test only, a Meaning test in which the target words were presented in their original context.

Results

These are given in Table 11.3. A comparison with the results of the Meaning tests administered in Experiments I and II shows that, obviously, performance was (proportionally) higher when the target words were given in context (Experiment III) than when given in isolation (Experiments I and II). More importantly, it was found that retention in the MC condition was substantially and significantly higher than in the Synonym condition.

2.6. Experiment IV

Subjects in this experiment were the students of two grade 9 classes (around 15 years old), all Ss being native speakers of Dutch. The students in one class did not know they would later be tested on their knowledge of the twelve target words (incidental learning task). The students in the other class, however, were told in advance that they would later be tested on their knowledge of the target words (intentional learning task). The cues, in the margin, given in the Synonym and MC conditions were identical to the cues used in Experiment II, with one exception: in the MC condition not four alternatives were given (as in Experiment II), but two alternatives only. Hence, there was one distractor only, and not three. Two Meaning post-tests were administered. In the first
Table 11.4  Experiment IV: retention of meanings by group and test

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Test 1 isolated</th>
<th>Test 2 in context</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Syn.</td>
<td>MC</td>
</tr>
<tr>
<td>Incidental</td>
<td>24</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Intentional</td>
<td>28</td>
<td>3.2</td>
<td>3.2</td>
</tr>
</tbody>
</table>

* A significant difference was found between MC and Syn. on Test 2 in the incidental group.

Results

These are given in Table 11.4. For the incidental group there was no significant difference between conditions on the test which presented target words in isolation. However, on the test which presented target words in context, the retention in the MC condition was significantly higher than in the Synonym condition.

Naturally, the retention scores of Ss in the intentional learning group was much higher than the retention scores of Ss in the incidental learning group. More importantly, however, there were no significant condition effects among the intentional Ss, on either test. Thus, as predicted in the design section, the effect of rehearsal and memorising techniques completely washes out the difference between the processing of given and inferred meanings.

2.7. Experiment V

Ss in this experiment were 35 adult non-native learners of Dutch. After Ss had served in the incidental learning situation and had completed two unexpected Meaning tests (target words in isolation and in context), they were given a 15-minute break. We then returned the text to them and told them that they would be, once again, tested on their knowledge of the words cued in the margin. They were given six minutes to study these words. Then we administered the two Meaning tests again, and finally the Form test.

Table 11.5  Retention of word meanings and word forms by task and condition (max. = 6; N = 35)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Incidental tests</th>
<th>Intentional tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Meaning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Isolated</td>
<td>In context</td>
</tr>
<tr>
<td>Synonym</td>
<td>0.7</td>
<td>1.6</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>1.6*</td>
<td>2.5**</td>
</tr>
</tbody>
</table>

A significant difference between conditions was found (*) on the incidental test with isolated target words, and (**) on the incidental test with target words in context.

Results

These are given in Table 11.5. On both incidental tests, the retention of the inferred meanings (Synonym) was higher than the retention of the given meanings. As in Experiment IV, performance on tests under intentional learning circumstances in Experiment V was substantially higher than performance on tests under incidental circumstances, while the differences between the MC and Synonym conditions disappeared. Performance on the Form test did not yield significant differences between conditions either.

2.8. Appeal of the distractors in the MC procedure

In order to assess the appeal of the distractors used in the present experiments, we counted how often subjects had selected and checked a wrong alternative (distractor) of those target words that were explained by means of an MC cue in the text margin. Note that in Experiments I, II and III (three distractors) the probability of selecting the correct meaning by chance was 25 per cent, while in Experiments IV and V (one distractor) this probability was 50 per cent. From Table 11.6, it can be seen that particularly in Experiments I and III, with non-native subjects, the percentage of wrong choices is alarmingly high.

3. GENERAL DISCUSSION AND CONCLUSIONS

Before stating our conclusions, we would like to draw the reader's attention to a plain but important fact: the retention of word
Wrong responses in text margin in MC condition

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Ss</th>
<th>Chance level in %</th>
<th>Wrong responses in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Non-natives</td>
<td>25</td>
<td>44</td>
</tr>
<tr>
<td>II</td>
<td>Natives</td>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>III</td>
<td>Non-natives</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>IV</td>
<td>Natives</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>V</td>
<td>Non-natives</td>
<td>50</td>
<td>8</td>
</tr>
</tbody>
</table>

meanings in a true incidental learning task is very low indeed. The chance that L2 learners/readers remember the meaning of a word, occurring only once in a text which is being read for its content, is very small. Of course, by stating this fact, we do not contend that reading for information is useless for the expansion of the learner’s vocabulary. Indeed, it has been convincingly argued that the enormous expansion of L1 vocabulary among children in elementary and secondary school can only be explained by a hardly visible incremental process, resulting from ‘normal’ reading-for-information, and that this expansion cannot be the result of committing lists of vocabulary items to memory and rehearsing these (Nagy, Anderson and Herman, 1987). Thus, while acknowledging the usefulness of reading for information even for vocabulary acquisition, one has to bear in mind that, with this procedure, it will only occasionally be the case that the meaning of a word can be learned fully upon a single occurrence.

After this preliminary observation, what conclusion can be drawn from the results of these five experiments? Are inferred meanings better retained than given meanings (first research objective), and to what extent can the MC procedure compensate for the inherent drawbacks of the inferring method? We will summarise our findings and conclusions in four points.

(a) If L2 learners/readers read an L2 text for comprehension of its content, and not with the intention to expand their L2 vocabulary, then they are more likely to remember the form and meaning of an unknown word in the text when they have inferred its meaning by themselves (high mental effort), than when the meaning has been given to them (low mental effort). Evidence for this conclusion stems from Experiments III, IV and V, using a within-subject design. In Experiments I and II, using a (weaker) independent-groups design, no evidence for this proposition was found.

(b) L2 learners/readers are more likely to infer an incorrect meaning of an unknown L2 word in an L2 text when no cue as to its meaning has been given than when a cue has been given (Experiments I and II).

(c) On the basis of (a) and (b), we argue that the discussion in foreign-language pedagogy should not focus on the question of whether it is better to give the meaning of an unknown word than let learners infer word meanings themselves without a cue (‘giving versus guessing’). The discussion should rather focus on the question of which cue procedures are most effective (e.g. translation into L1, synonym in L2, concise sample sentence, multiple choice, separately or combined).

(d) Although the multiple-choice procedure, as an instantiation of an inferring method, has been shown to have a higher retention effect than the synonym procedure (conclusion (a)), it was also found to be inherently error prone: in all experiments there were some subjects for whom some distractors were too appealing (Table 11.6). From this finding we conclude that the MC procedure should only be used when a teacher is present. First of all, the teacher is best capable of deciding which distractor(s) would best suit his/her students, i.e. which distractors are not too plausible nor too implausible for the students. Second, it is the teacher who can correct immediately, should students choose a distractor nonetheless. Hence, the MC procedure, as any other open technique offering L2 learners/readers a cue as to a word’s meaning, seems inappropriate when students are working on their own. Thus, for reading L2 texts at home (unguided L2 learning), the classical ‘closed’ (glossing) procedures, such as giving a synonym or translation, seem to constitute safer cue procedures, albeit possibly less effective, than ‘open’ procedures, such as giving concise context sample sentences or multiple-choice questions.

Notes

1. I would like to thank two graduate students at Leyden University, Ans Mahieu and Inge Wellecomme, who helped me with administering and scoring the tests in Experiments I and IV. I am also grateful to the Stedelijke Scholengemeenschap Middelburg for allowing us to test two classes (Experiment IV).
2. There is no room here to present all details concerning materials, subjects, procedures, results, statistical analyses, etc. However, a detailed report on these five experiments is available on request from the author: Dr Jan H. Hulstijn, Applied Linguistics Department, Free University, PO Box 7161, 1007 MC Amsterdam, The Netherlands.

References


