Learning Words from Dictionaries and Context: a Replication

James Ronald

(Received on May 18, 1999)

Abstract

Two means by which second language learners encounter and learn new words are through reading L2 texts and by looking up words in monolingual learners' dictionaries. How the word knowledge gained via these means differs is not fully understood. This paper reports a replication of an experiment by U. Fischer (1994) investigating this question, although differing from Fischer's experiment in two main ways: the subjects' L1 and the source material. The results of the study did confirm the greater accuracy of subjects using dictionary definitions, but they also highlighted a number of major design faults in the experimental methods employed.

Introduction

This paper is a report of a replication of an experiment by Ute Fischer (1994), comparing the learning and use of English words from monolingual dictionary entries with words encountered in the text of a novel. The procedure for the experiment was basically the same as Fischer's, as was the coding of the data. These are described below.

The subjects and materials differed in some respects from those in Fischer's experiment. The subjects were Japanese university students rather than German high school students. The materials were entries from *Collins COBUILD English Dictionary* (Sinclair et al., 1995) rather than the *Oxford Advanced Learner's Dictionary* (Hornby et al., 1980) and a passage from Sidney Sheldon's *Memories of Midnight* (1990) rather than from John Fowles' novel *The Collector* (1981). In addition, the target words in this experiment were all adjec-

tives rather than a mixture of nouns, verbs, and adjectives.

The goal of the experiment reported in this paper is to investigate the same questions as Fischer:

- · Which of the sources of information was most helpful to learners in their comprehension and use of the target words.
- · How subjects with both sets of materials (dictionary entries and text) used the materials.
- · In using the target words, what strategies were employed by the subjects from the different groups.

Method

Sixty-nine second year English major Japanese university students were recruited as subjects. All of these students were native speakers of Japanese. They were intermediate to advanced learners who had received about seven years of formal instruction in English. The subjects were between 19 and 21 years old. Seventeen of the subjects received dictionary entries of unfamiliar words, seventeen received a narration in which the same target words were embedded, and eighteen received both dictionary entries and text. The seventeen subjects who received the text but no information about the target words served as controls.

Stimulus material

Target words   There were 12 target words, all adjectives (*acrid, armed, budding, diverting, invincible, noisome, petrified, relentless, stagnant, swollen, unerring, unnerving*). The words were selected on the basis of their not being included in *Eigo Tango 2001* (English Words 2001), a semi-official list of words used by students preparing for university entrance examinations (Uryu et al., 1993). The target words were printed on a sheet of paper in the order

that they occurred in the text.

Learning materials   The study involved two sources of information about the target words: their dictionary entries and part of a story in which they occur.   The dictionary entries were taken from the *Collins COBUILD English Dictionary*, 2nd Edition (1995).   None of the subjects were regular users of a monolingual learner's dictionary.   *COBUILD* was chosen because the students had some familiarity with the layout and style of its entries; it had been used a few weeks prior to the study to introduce the students to monolingual dictionaries.   Only two of the definitions contained information about the context in which the words were used (formal, old-fashioned), although a total of eleven of the entries indicated semantic restrictions (*If you describe an army or sports team as invincible... Stagnant water is...*).   For eleven of the words, definitions were followed by sentences or phrases that illustrated word use.   The study also used utilised a passage adapted from Sidney Sheldon's *Memories of Midnight* (1990).   The part chosen was adapted to only contain simple syntactic constructions and basic vocabulary.   All of the words except the target words could be found in *Eigo Tango 2001* (Uryu et al., 1993).   Five of the target words were already in the original text, while seven were added to the text.   A second version of the text, for the control group, was obtained by deleting the target words.

Procedure

Two classes of university students majoring in English participated in the study.   Each test lasted a total of 90 minutes and consisted of three successive parts: (1) word explanation, (2) acquisition phase, (3) word explanation.   As there was insufficient time to conduct the whole test in one 90-minute class, (3) was conducted one week later than (1) and (2).   This also allowed for a more meaningful interval between acquisition and recall.

Word explanation   The test was given twice: once just before the acquisition phase and then, unexpected by the subjects, again one week after.   On both occasions subjects were given the list of target words and had to state what each meant.   The instructions, written in Japanese, emphasised that they could respond in either Japanese or English.   The pretest showed which words were already known by the subjects, and the posttest indicated how well they could explain the words they had encountered during the acquisition phase.

Acquisition phase   The study involved three experimental conditions and one control condition.   One experimental group received the sheet of dictionary entries for the target words, another group was given the text, and the third group was given both the dictionary entries and the text.   Subjects in the control group received the version of the text in which the target words had been deleted, and were required to write a summary of the story in Japanese.   Subjects in the experimental conditions were asked to read carefully through the information they had received and then to use each target word in a sentence.   After about 45 minutes all subjects indicated that they had completed the required task.   The subjects were then asked to write a Japanese translation for each of their English sentences.   Prior to this instruction there had been no mention of the translation task.

Coding the data

Evaluation of the summaries of the control subjects   Two raters independently made up a list of important points that they felt should be included in a summary of the story.   Eight facts were considered important by both raters.   Two raters then evaluated how many facts from this list were stated in the summaries of the students.   Each fact was counted as one point; partially recorded facts were counted as 0.5 points.   Although the raters agreed

only 15% of the time, they differed in their judgment by one point or less for 71% of the remaining protocols. All disagreements in the evaluations were resolved through discussion.

Coding the English sentences Two native speakers of English independently rated the target word in each English sentence in the following way: whether the target word was used in an idiomatically meaningful way, whether its usage was questionable, or whether it was used in idiomatically unacceptable contexts. The sentences were alphabetised and did not contain any indication as to which experimental group the writer belonged to. Interrater reliability was .68; disagreements between raters were setted through discussion.

Coding the Japanese translations A native speaker of Japanese evaluated how well the subjects' translations matched a standard, namely the Japanese equivalent of a target word as stated in bilingual dictionaries: *Kenkyusha's English–Japanese Dictionary for the General Reader* (Matsuda et al., 1992) and *An Encyclopedic Supplement to the Dictionary for the General Reader* (Matsuda et al., 1994). It was judged whether a translation was a match, a near-match, a far-match, or a no-match. A monolingual dictionary, *Kojien Dainihan Hoteiban* (Shinmura et al., 1976), was also used to determine the adequacy of subjects' translations. After the rater had completed the coding, a second rater, also a native speaker of Japanese, was asked to check the judgments of the first speaker. The raters agreed in 75% of the instances. Disagreements were resolved through discussion.

If a translation was equivalent to a standard, it was called a match. A translation that was not semantically related to the standard was a no-match. A translation was rated as a near-match if it was a superordinate of the standard; for example 汚い 水 ("kitanai mizu"— *dirty water*) as translation for stagnant (*water*). If the meaning of the standard implied the meaning of the translation, or vice versa, and their semantic relation could not be characterised in

terms of hyponymy, then the translation was classified as a far-match; for example, noisome rendered as ひどくつまらない ("hidoku tsumaranai"— *terribly boring*).

Coding the strategies.　The same procedure was employed as for the coding of the Japanese translations.　The main rater coded the strategies that the students in the experimental conditions seemed to have employed, and the second rater checked the ratings.　Subjects' strategies were inferred from their translations of the target words and the learning materials they had seen.　The raters agreed 80% of the time.　All disagreements were resolved through discussion.

Dictionary group.　The strategies identified in subjects' work in this condition were based on Fischer's and are listed in Table 1.　The occurrence of a substitution strategy was noted whenever the translation of a target word was based on all or part of the definitional information.　Based on this distinction, two classes of substitutions — complete string substitutions and substring substitutions — were observed.

Responses in which subjects directly incorporated information from either definitions or illustrative phrases were termed as copying or modelling. Copying was coded when an English sentence involved part of the definition or an example.　Beside using information verbatim, subjects also modelled their sentences after an example or the definition.　A false positive was coded as a strategy when the translation (e.g. うるさい —"urusai"— 'noisy', for *noisome*) was arrived at by the mistaken linking of the target word with a similar known English word.　This type of strategy is different from the one in Fischer; as there are no real English-Japanese cognates there was no confusion in this area.　On the other hand, confusion between similar sounding English words was not unusual so this became the 'False Positive' category.　Unfinished substitution errors referred to English that failed to

include the target word but included part of its definition instead.

Table 1   Comprehension strategies of the subjects in the dictionary group.

Strategy	Description	Example	Dictionary entry
Complete string substitution	Translation utilizes complete definition	I was petrified because I saw a ghost.	If you are petrified, you are extremely frightened, perhaps so frightened that you cannot think or move. *I've always been petrified of being alone... Most people seem to be petrified of snakes.*
Substring substitution	Translation uses part of definition	After exercising a soccer for a long time I am always petrified.	
Modelling	Sentence is modelled after definition or example	She seem to be petrified of ghosts.	
Unfinished substitution error	Sentence includes part of the definition (but not the target word)	Our team will not *unbeatable* in the game.	=*unbeatable* (given as synonym in entry)
False positive	Translation is equivalent of word with typography similar to the target word	Since it has been 50's the earth is being influence by acrid rain. (*acrid* mistaken for *acid*)	

Text group.   Table 2 lists the strategies for the text group.   The English sentences written by subjects with the text as source were rated according to whether they incorporated the lexical context in which the target word occurred.   It was noted whether a target word's context was adopted verbatim or whether subjects' sentences were modelled on the sentences in which the target words occurred.   In addition, it was judged whether a translation suited only the lexical context of a target word or whether it also accorded with its schematic context.   If an English sentence did not incorporate the lexical context of a target word, it was judged whether the translation of the target word was consistent with all or only part of its schematic context. Translations that cohered with neither the lexical nor the schematic context

were coded separately as unexplainable.

Table 2. Comprehension strategies of the subjects in the text group.

Strategy	Description	Example	Text
L+S	Sentence incorporates lexical context and translation accords with lexical and schematic context	The docks stank of the stagnant mass of dead cats. (よどんだ—'yodonda' = stagnant)	The docks stank of the stagnant mass of dead cats and dogs.
L	Sentence incorporates lexical context and translation accords only with lexical context	She is an unerring typist. (間違いの ない—'machigai no nai' = without mistakes)	...Tony's boxing skills and unerring killer instinct.
S	Sentence and translation accord only with schematic context	He knows of the unerring way of guns.	
P	Sentence and translation accord only with part of the schematic context	I think he went on relentless exercise every day.	...the school itself was a relentless battle ground.

Mixed group. The coding system for the text- and dictionary group is summarized in Table 3. Translations that were literal or nonliteral translations were taken as evidence that the students had relied on the dictionary. Students were assumed to have focused on the text when they used a target word in a

Table 3. Comprehension strategies of subjects in the mixed group.

Strategy	Description	Example
Dictionary As source	Translation of definition, and/ or copying and modelling	I believe the soccer team is invincible.
Text as source	Sentence incorporates lexical context	The soccer game made me invincible.
Both sources	Translation accords with definition, and sentence similar to text: or two translations	The members think themselves invincible.

construction similar to its lexical context in the text. Students were judged to have considered both sources of information when they translated a target word in accordance with its definition and used it in an English sentence that was analogous to the text.

Results

Results of the pretest

With the exception of one word (*armed*), the percentage of subjects who gave an appropriate English synonym or Japanese translation to the target words in the pretest was very low. As can be seen in Table 4, most subjects were unfamiliar with the target words. As will be explained, the data of the few subjects who were familiar with a target word were excluded from the analyses that are reported below.

Results for the control group

55% of the control subjects mentioned 5 or more of the important facts of the narration. Of the remaining 8 subjects, 6 mentioned more than 50% of impor- tant points. This suggests that subjects, generally, had a fair understanding

Table 4. Percentage of subjects familiar with target words prior to the study.

Word	Dictionary group	Text group	Mixed group
acrid	0	0	0
armed	41	12	22
budding	0	0	0
diverting	0	0	0
invincible	0	0	0
noisome	0	0	0
petrified	0	0	0
relentless	0	0	0
stagnant	0	0	0
swollen	0	0	0
unerring	6	6	11
unnerving	0	0	0

of the story.

Adequacy of usage and comprehension of target words

Table 5 shows the percentages of appropriate and inappropriate uses of the target words in the English sentences that were obtained for each experimental group. As can be seen, subjects in the dictionary and text groups performed markedly better than the text group.

Table 5. Percentage of omissions, idiomatically incorrect, questionable, and correct uses for dictionary, text, and mixed groups.

Adequacy	Dictionary group	Text group	Mixed group
Omissions	1	1	2
Incorrect	10	25	7
Questionable	25	34	28
Correct	63	40	60

Table 6 summarizes how accurately subjects in each experimental condition translated the target words. For each subject, the number of matching, near-, and far-matching translations were collapsed into one score, and labelled "adequate translations". The type of information that was available to subjects influenced their comprehension. Subjects in the dictionary and the mixed groups gave, on average, 9.88 and 10.39 correct translations, respectively. The mean number of translations for the text group was

Table 6. Percentages of omission, no-, far-, near-, and matching translations for dictionary, text, and mixed groups.

Adequacy	Dictionary group	Text group	Mixed group
Omissions	1	1	2
No-match	17	63	11
Far-match	6	12	13
Near-match	29	10	23
Match	48	14	51

4.29.   For the English sentences, omissions and questionable or unaccept-able English uses were collapsed into one category — bad usage.   Adequate uses were good usages.   Subjects in the dictionary and the mixed groups gave, on average, 7.65 and 7.06 good English sentences, respectively, while the mean figure for the text group was 4.65 good sentences.
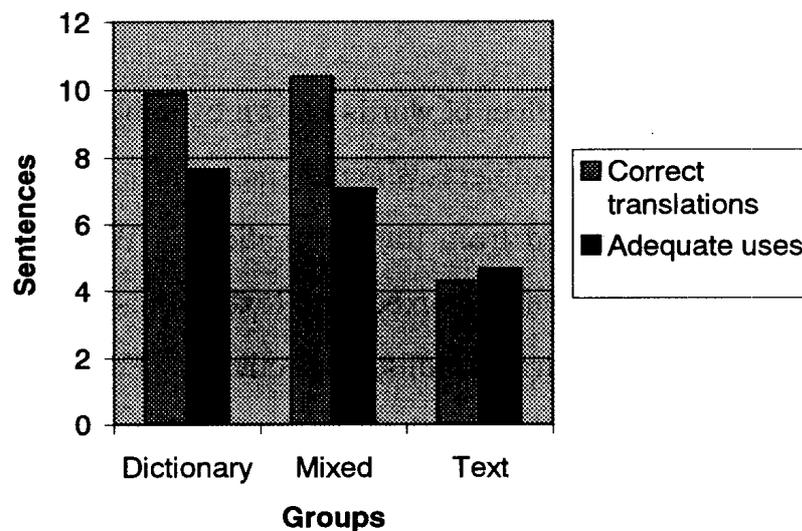

Figure 1.   Average numbers of correct translations and adequate English sentences per group.

A Oneway analysis of variance was conducted on the ratings of the Japanese sentences per word for every subject within each group.   For this analysis, matching, near-, and far-matching translations were collapsed into one score, as above, as "adequate translations", with omissions and no-matching transla-tions counted as "inadequate translations".   For the translations, a Oneway analysis of variance showed that there was a significant difference between the groups [F (2,49) = 47.38, p < .001].   Tukey tests confirmed that the text group is performing worse than the two other groups, but that there was no difference between the dictionary and mixed groups.

A Oneway analysis of variance on the ratings of the English sentences per word was conducted for every subject within each group.   Similarly, in this

analysis omissions and questionable or unacceptable English uses were collapsed into one category —"bad usage". Adequate uses were "good usages". This analysis of the English sentences showed a significant difference between the groups [F (2,49) = 7.63, p < .001]. For the English sentences as well, Tukey tests confirmed that the text group is performing worse than the two other groups, but that there was no difference between the dictionary and mixed groups.

In order to eliminate the effect of words known by subjects prior to exposure to the materials, the few subjects who did know one or more of the words in the pretest were excluded from the calculations. With these revised groups, subjects in the dictionary, mixed, and text groups gave, on average, 9.78, 10.14, and 4.21 correct translations, respectively. For the English sentences, subjects in the dictionary and the mixed groups gave, on average, 8.55 and 6.79 good English sentences, respectively, while the mean figure for the text group was 4.21 good sentences, as shown in Figure 2.

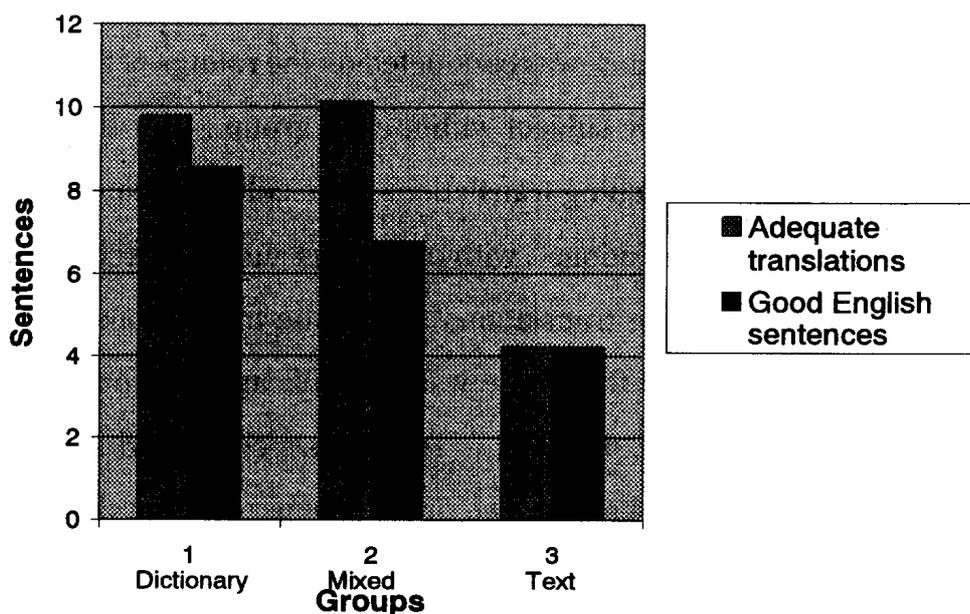When a Oneway analysis of variance was performed on these new groups,

Figure 2. Average numbers of adequate translations and good English sentences for the revised groups.

there is still a significant difference between the groups' results, both for the translations, [F (2,34) = 27.95, p <. 001], and for the English sentences: [F (2,34) = 10.21, p < .001]. Tukey tests with the revised groups' results confirm that the text group performed worst of the three and that there was no significant difference between the dictionary group and the mixed group.

Students' strategies

In the dictionary group, 87% of subjects' responses were accounted for by the observed strategies, with subjects most commonly adhering to a substitution strategy. More importantly, subjects focused 50% of the time on all available definitional information. Substring substitutions, which indicate an inadequate understanding of a definition, accounted for 35%. Copying, which always co-occurred with other strategies, was noted in 7%, and modelling was observed in 4% of the sentences subjects wrote.

The strategies that were discerned for the text group can explain 35% of the subjects' responses. Subjects in the text group use both the lexical and the schematic context of a word to infer its meaning but this was evident in only 10% of sentences. In 12% of their sentences, subjects preserved the sequence of a target word and adjacent words, while only 3% of their sentences accorded with the schematic context of the target words. A total of 64% of sentences could not be assigned to any of these categories.

Overall, 98% of the responses of the subjects in the mixed group could be explained in terms of the strategies identified in the coding scheme. Of their responses, 80% could only be traced to information in dictionary entries. They focused on the text only 2% of the time, and 16% of their responses were based on both dictionary and text. For the sentences in the mixed group where they used the dictionary entries, the frequencies of complete string substitution and substring substitution are comparable to the ones observed for sub-

jects in the dictionary-only group: 54% and 35% respectively.

Discussion and conclusion

The discussion will focus on two matters relating to this experiment:

i) How much the experiment provides answers to two of the questions posed in the Introduction.

ii) How the differences in the results of Fischer's experiment and this replication may be attributable to the differences in the materials used in the two studies.

i) If we return to the questions in the Introduction, we can consider to what degree the research described above may provide answers:

"Which source of information was most helpful to learners?" The two groups with access to dictionary entries clearly produced the largest number of accurate uses of the target words in the task.

"What did subjects in the mixed group do with the two sets of materials?" In the mixed group, as with Fischer, subjects appeared to rely largely on the dictionary entries alone, and their results showed no significant differences from those of the dictionary group.

The greatest difference between the results in Fischer's experiment and this replication is in the production of correct English sentences. While for Fischer, the results of the three groups were very similar, my results showed the text group to be markedly weaker than the other two groups. Two differing features of the materials provided for the subjects in the two experiments may help explain these differences.

a) For the text used by Fischer, all the target words were inserted into the text and so could be presented in a clear, comprehensible, and fairly typical context. In my text, five of the words were already a natural part of the selected text. In any single work of fiction, there is no guarantee

that each word will be in a clear, comprehensible, or typical context. This may be one reason for the relatively weak performance of the text group in my experiment.

b) Another factor, this time accounting for the relatively strong performance of the two groups with dictionary entries in my experiment, may be the difference between the two sets of dictionary entries. Two major differences are presence or absence of indications of semantic restriction for the words and the absence or presence of example sentences or phrases in the dictionary entries. In my set of twelve entries, eleven provided clear indications of semantic restrictions (*Stagnant water is... An acrid smell or taste is...*) and eleven also included at least one, and often two or three, long example phrases or sentences. In Fischer's case, semantic restrictions are harder to identify and there are fewer of them: just two or three out of twelve definitions. Seven of the definitions were followed by an example phrase, but these were usually only two or three words long.

Conclusion

Finally, in undertaking this replication, what has become clear is that although the question of how language learners can benefit from different sources of information is an important one, neither the original experiment by Fischer (1994) nor this replication offers any reliable answers. The results described and analysed above depend to a very large degree upon a number of factors that are not reliable; texts that may contain untypical examples of word use and raters' subjective evaluations of correctness and acceptability. In addition, what might seem to be the most reliable and replicable aspect of the experiment, dictionary entries, revealed large discrepancies between the type

and amount of information provided for different words in the two dictionaries used.

References

Fischer, U. (1994) Learning words from context and dictionaries: an experimental comparison. *Applied psycholinguistics*, 15, 551–574.

Fowles, J. (1981) *The Collector.*

Hornby, A. S. et al. (1980) *Oxford advanced learner's dictionary of current English*, 3rd edition. Oxford: Oxford University Press.

Matsuda et al. (1992). *Kenkyusha's English–Japanese dictionary for the general reader.* Tokyo: Kenkyusha.

Matsuda et al. (1994). *An encyclopedic supplement to the dictionary for the general reader.* Tokyo: Kenkyusha.

Sheldon, S. (1990) *Memories of midnight.* New York: Morrow.

Shinmura et al. (1976) *Kojien dainihan hoteiban* (2nd edition): Tokyo: Iwanami Shoten.

Sinclair, J. et al. (1995) *Collins COBUILD English dictionary*, 2nd edition. London: HarperCollins.

Uryu, Y. et al. (1993) *Eigo tango 2001* (English words 2001). Tokyo: Kawai Publishing.