Efficiency of Information Retrieval from the Electronic and the Printed Versions of a Bilingual Dictionary

SHIZUKA Tetsuhito
Kansai University

This paper reports on a study that compared electronic and printed dictionaries in terms of empirical and perceived efficiency of meaning and example retrieval. Seventy-seven university students took a speed test that measured efficiency with which word meanings (Part 1) and examples (Part 2) were accessed. The participants used electronic or printed versions of the same English-Japanese dictionary. A post-experiment survey explored participants' perception of the two dictionary types. Participants' pre-experiment familiarity with electronic and printed dictionaries was also examined. Multiple regression models were fit in search of meaningful relationships between the variables. The results indicated that (1) in identifying word meanings, e-dictionaries were markedly more efficient; (2) this advantage was multiplied by the users' familiarity with e-dictionaries; (3) in accessing examples, there was no significant difference between the two dictionary types; and (4) participants overwhelmingly preferred e-dictionaries. The paper concludes with an argument that in light of much less degree of reluctance to use e-dictionaries, the electronic-printed gap in real-life use frequency is expected to be larger than was observed by the speed test in this study.

1. Introduction

Few dispute that vocabulary expansion is crucial for successful language learning and that using dictionaries is one, if not the most, important strategy for that purpose (Barnet, 1989; Knight, 1994; Nation, 2001). Consulting dictionaries by flipping through pages and struggling with reading the fine print, however, has traditionally been perceived as an "arduous" task (Keller, 1987, p. 17). The advent of electronic dictionaries (e-dictionaries) has completely changed that image. Lexical information embedded in reading materials using hypertext technology naturally allows much quicker word search (Bhatia, 1991; Koga, 1996) and consequently results in more frequent consultation (Aust, Kelley, & Roby, 1993). A still further development is the recent emergence of pocket-sized electronic versions of reputable printed dictionaries. Unlike their predecessors that displayed nothing but sketchy definitions, these "full-contents" pocket e-dictionaries contain all the information found in their corresponding printed ones minus illustrations.

CASIO Computer Co. estimates the current number of pocket e-dictionaries used in Japan to be around 4 million. According to a survey conducted by the corporation in 2000, 70 percent of the respondents--400 high school students in Tokyo--owned pocket e-dictionaries. Takada's (2002) survey found that 60 percent of the respondents--79 high school students in Kyoto--were e-dictionary users and 69 percent of the non-users were aspiring users. Surveys carried out by CASIO in 2001 and 2002 on university students revealed that the proportion of e-dictionary users grew almost fourfold in that one year, from 10 to 38 percent.

Despite their popularity, though, how exactly these digital devices compare with their printed counterparts as language-learning tools is still mostly an open question. Primarily because "full-contents" pocket e-dictionaries are relatively new inventions, the number of studies that focused on their potentials is still small. The results of those that did, however, are mixed.

Tsuchimochi (2002, as cited in Yanase, 2002) measured how many words university students could look up in a speeded situation, on electronic and printed versions of *Genius English-Japanese Dictionary*. She had 46 subjects look up as many words as possible in four three-minute trials. On average, the number of words accessed on e-dictionaries in three minutes was 15.40, and that on p-dictionaries was 7.33 ($t = 17.38$, $p < .01$). Tsuchimochi then lent e-dictionaries to ten subjects for a week and took surveys on voluntary use frequencies during that period. Most subjects reported that they had used e-dictionaries "incomparably more frequently" than p-dictionaries.

Koyama and Takeuchi (2001a, 2001b, 2002a, 2002b) have been conducting a series of studies on electronic-printed comparability, from which emerges a picture not totally rosy for e-dictionaries. Their first study (Koyama & Takeuchi, 2001a) on high school and junior college students' consultation behaviors did not find any significant differences in search frequency, time spent on search, or retention of searched words. Noted also was the tendency for users' familiarity with each dictionary type to affect their preference. Their second study (Koyama & Takeuchi 2001b) based on the think-aloud protocols produced by four university students revealed, among other things, e-dictionary users' frustration with the limited amount of information viewable at one time. The third study (Koyama & Takeuchi 2002a) indicated that, though search time was not significantly different across dictionary type, search on p-dictionaries resulted in better retention of searched words. Similar results were reported by Koyama and Takeuchi (2002b) as well.

The present study attempted to shed further light to the still equivocal nature of pocket e-dictionaries. Specifically, it addressed the following questions:

(1) How efficient are e-dictionaries, relative to p-dictionaries, in simple word search?

(2) How efficient are e-dictionaries, relative to p-dictionaries, in example search?

(3) How are word- and example-search efficiencies with p- and e-dictionaries affected by users' familiarity with each dictionary type?

(4) How do learners perceive relative efficiencies of e- and p-dictionaries in information retrieval and general effectiveness as learning tools?

2. Method

2.1. Participants

The participants were two classes of Japanese university 1st year students enrolled in an English as a foreign language course. After removing data by those who failed to complete one or more of the tasks, 77 were retained for analysis. Though in reality data collection was conducted separately in two classes consisting of approximately 40 students each, descriptions in the procedure section below will be given as if they were one large class of approximately 80 students.

2.2. Instrumentation

2.2.1. Dictionaries

Printed and electronic versions of the same dictionary needed to be prepared. Chosen for the present study was *Genius English-Japanese Dictionary*, 2nd Edition (Taishukan). The original printed version was available from Taishukan and an electronic version was contained in XD-S1200, produced by CASIO.

2.2.2. Information Retrieval Speed Test:

Two equivalent forms (Forms A and B), each consisting of Part 1 (10 items) and Part 2 (5 items) were developed.

2.2.2.1. Part 1

Objective

Part 1 items were developed to estimate facility with which dictionary users look up unknown words about which polysemy was not a problem, and identify their meanings (in the form of L1 equivalents). This corresponded to a real life situation where, for example, a Japanese university student preparing for the next English class encounters an unknown word in the textbook, looks it up in a bilingual dictionary, locates the headword, identifies its meaning in L1, jots it down, closes the dictionary, and proceeds. Informal observation of Japanese university students indicates that this type of procedure that does not include careful examination of usages or detailed processing of examples is the most common among their dictionary-using behaviors.

Hence the ten items in Part 1 required the test-taker simply to look up the designated words in the dictionary and choose the most appropriate L1 equivalent from among four options provided, as follows:

Look up the following words in your Genius EJ Dictionary and choose the most appropriate Japanese equivalents. (Originally in Japanese)

1. scam A *kabunushi* (shareholder) B *komon* (advisor)
C *sagi* (scam) D *shintaku* (trust)

Item Format

Although looking up words without context was not authentic, target words were presented in isolation because that would better suit the purpose of this research. The variable in question was the speed with which the target words could be located and their meanings identified. Context of course may, or may not, facilitate and swiften this process, but that facilitation is likely to occur to different extents across persons, resulting in unwanted contamination. Not presenting context information to any word was deemed better for measuring the mechanical speed.

Target Word Selection

The item format described above dictates the type of target words suited to this part: the words needed to be those whose meanings depended little on context. This restriction led us to decide on nouns with relatively low frequency of occurrence.

All the selected words were of or below *Genius* Rank C, meaning that they were not among the most "important" 6000 words (see Table 1). It was expected that very few, if any, of them were known to the participants. To make it doubly certain that the participants actually look up the words, it was emphasized that since the objective was not to test their vocabulary knowledge but to examine user-friendliness of different dictionary types, they had to look up all the target words even when some of them looked familiar.

They were also controlled for length, between forms, since the number of characters in a word has a direct impact on the time duration required to locate the words in a dictionary, especially in a printed one. In both forms, the words were arranged so that they gradually increased in length as the test-taker proceeded in the test. The lengths of the words in the two forms with the same item number (e.g., Item 5 in Form A and Item 5 in Form B) were made exactly the same.

Table 1. Target words for Part 1, Form A and Form B

Form A	scam; brink; obesity; toddler; smallpox; incision; ceasefire; hypocrite; appraisal; respiration
Form B	whim; havoc; curator; refugee; epidemic; emission; nutrition; assailant; deterrent; miscarriage

Distractors and the Correct Option

The distractors were all L1 words corresponding to some noun concept. They were written so that when the meaning of the target words were not known, all of them may look plausible, but once the target word meaning is identified in the dictionary, all of them were unmistakably wrong. The correct option was simply the first L1 equivalent that appeared under the headword in *Genius English Japanese Dictionary*. This assured that locating the word in the dictionary almost automatically led to choosing the correct option.

Problem Solution Order

The participants were also instructed that they should respond to the items strictly in the designated order and that skipping was not allowed. This control for the order of item solution was to ward off possible contamination from test-taking strategies. If all the participants look up the target words in the same order, the variable in question, the speed with which the words could be located, was expected to be tapped more precisely than otherwise.

2.2.2.2. Part 2

Objective

Part 2 was designed to tap participants' speed with which they locate information regarding the target words' usage. This part was intended to concern such dictionary-using phases as locating the headword, identifying relevant sections, and examining information concerning the usage and/or example sentence/phrases.

This phase was of interest mainly for two reasons. One was that processing examples was undoubtedly crucial for vocabulary learning, particularly so if the words were to be used for production. The other reason, the one more important when discussing the effectiveness of portable e-dictionaries relative to p-dictionaries, was that most of the current ones present information in a hierarchical manner.

Hierarchical Structure

Due to limitation of their screen sizes (typically, 5-7 by 10-12 centimeters), portable e-dictionaries present information on two-levels. When the spelling of a word is input, the first screen (Level One) will only display the summary of the available information. Shown at this level are pronunciation, word-class(es), minimal explanation in L1, L1 equivalents, and derivatives when there are some. Examples are not displayed at this level.

When examples are available for a particular meaning, a button-like label 用例 ("example") is shown beside that section. When that label is hi-lighted and a button pushed, illustrative

sentence(s) for that meaning will appear (Level Two).

In the case of words with more than one meaning category, with examples accompanying each category, the examples for a particular meaning are stored in a subsection under that particular meaning. Therefore, in order to locate examples containing the target word used in a particular way, the user needs first to locate the right numbered-section and then push the 用例 button in that section. When the user reaches a wrong subsection, he/she needs to go back up to the Level One screen, and then try another section before locating the correct one.

Hence, locating relevant examples on an e-dictionary requires a behavior very different from that needed for example search on a p-dictionary, where all available information is presented at one and the same level, at a glance.

Item Type

As stated above, of interest was the ease with which participants locate examples in which the target word is used in a particular way. Unlike when the goal was identifying the meaning in L1, however, examples themselves did not readily allow creating multiple-choice options. For this reason, the item type adopted for Part 2 was as follows.

Look up the following headwords in Genius EJD, locate the designated examples, and copy the words used in the places of the blanks. (Originally in Japanese)
1. take <人が><ある行動>をする take legal ( ) 訴訟を起こす

Provided as stimuli were (1) a target word, (2) one of its meanings in L1, (3) an example with a one-word blank, and (4) the translation of the example. Required as a response was to fill in the blank. The examples with their translations provided in the test were exact copies of those appeared in *Genius English Japanese Dictionary*. Hence, the subject needed to (1) access the headword, (2) identify the section for the specified meaning, and (3) locate the designated example.

Target Word Selection

The type of words most suited to this item type was decided to be high frequency verbs with multiple meanings with numerous examples. Chosen were ten (five for each form) verbs, which must have been more than familiar to all the participants. Though it was not possible to exactly match the number of characters in the target words between forms, the mean lengths were comparable to each other (3.6 for Form A and 3.8 for Form B). Table 2 shows the number of meaning categories under each headword. It is apparent that the target words in this part had many times more meanings and more examples than those used in Part 1. Participants needed to search

through all these sections to locate the right one, and then search through multiple examples to locate the designated one. The instruction on the order of item solution was given with regard to this part as well; the items were to be responded to strictly in the order they appeared.

Table 2. Target words and examples for Part 2, Form A

Form A	Target word	Searched example	Number of MC*
Item 1	take	take legal (action)	49
Item 2	bring	bring the fire under (control)	8
Item 3	have	I am not having such (conduct) here.	28
Item 4	go	Everything must go. No early (bird).	37
Item 5	keep	keep the wine (out) for her	26
		Mean	29.6
Form B	Target word	Searched example	Number of MC*
Item 1	give	The car gave a (jolt).	25
Item 2	put	The news put me to (silence).	13
Item 3	make	(Charm) makes a salesman.	37
Item 4	get	Let's get (going)!	27
Item 5	come	Don't come the (bully) with me.	19
		Mean	24.2

*MC: meaning categories

2.2.3. English Proficiency Measure: A 122-item c-test

As a general English proficiency measure, a 122-item c-test, developed by Professor Amma Kazuo at Tamagawa University, was used.

2.2.4. Questionnaire

An 8-item survey was prepared (see Table 3).

Table 3. Post-experiment survey questions

Q1	How frequently do you use [have you used] printed English-Japanese dictionaries?
Q2	How frequently do you use [have you used] electronic English-Japanese dictionaries?
Q3	To what extent was it difficult or easy to simply locate (or make appear) specified entries?
Q4	To what extent was it easy to locate specified examples?
Q5	To what extent were located entries easily viewable or eye-friendly?
Q6	If you owned one of these, how willingly would you use it?
Q7	If you owned one of these, how useful would it be for learning English?
Q8	How good was the overall impression?

Questions 1 and 2 (referred to as *background questions* hereafter) were for exploring participants' familiarity with e- and p-dictionaries. These two questions were responded to by

rating on a 6-point scale (6 being "Every day" and 1 being "Never"). Questions 3-8 (*impression questions*) concerned participants' perceptions of printed and electronic dictionaries after using them for the speed test described above. These six questions each elicited two ratings, one for p-dictionaries and the other for e-dictionaries, on a 5-point scale (5 being "Very" and 1 being "Not at all").

2.3. Data Collection Procedure

Data were collected in such a way that each subject experienced using both types of dictionaries (see Table 4).

Table 4. Data collection procedure

	Group E-P (n=38)	Group P-E (n=39)
Form A (7.5 min)	Part 1 electronic (2.5 min)	Part 1 printed (2.5min)
	Part 2 electronic (5 min)	Part 2 printed (5 min)
Form B (7.5 min)	Part 1 printed (2.5 min)	Part 1 electronic (2.5 min)
	Part 2 printed (5 min)	Part 2 electronic (5 min)
	Questionnaire (5 min)	
	C-test (30+ min)	

Participants were first divided into two groups of approximately the same size, based on the rows of classroom seats they were sitting in. In the first phase, both groups tried Form A, half of them using p-dictionaries and the other half with e-dictionaries. Time allowance was 2.5 minutes for Part 1 and 5 minutes for Part 2. Then, the two groups swapped dictionaries and tried Form B, this time the first group using e-dictionaries and the second group using p-dictionaries. Time allowance was the same as that for Form A. After that, both groups answered survey questions for about 5 minutes. Finally, they took the c-test in about 30 minutes. The group who used e-dictionaries before p-dictionaries will be referred to as Group E-P (n = 38), and those who used p-dictionaries first will be termed Group P-E (n = 39).

3. Results

3.1. Pre-experiment Group Comparability

The c-tests results and responses to background questions were examined first to check the comparability of Groups E-P and P-E.

3.1.1. C-Test

The c-test number correct means for Group E-P and Group P-E were 77.71 and 80.97, respectively, and the standard deviations 26.14 and 14.39, respectively. Since an F-test revealed a

significant difference between the variances of the two groups ($F = 3.24$, $p < .01$), a t-test that did not assume equal variances was run. The result indicated that the two means did not differ significantly ($t = -0.47$, n.s.). Thus, there was no evidence that the two groups differed in terms of proficiency. Coefficient alphas were very high: .977 for Group E-P, .919 for Group P-E, and .992 for the total group. These values indicate that the test differentiated this particular group of participants very effectively and reliably in terms of the test score.

3.1.2. Background Questions

The responses to the background questions are summarized in Tables 5 and 6. The medians for Question 1 (p-dictionary familiarity) were 4 for Group E-P and 3 for Group P-E; and those for Question 2 (e-dictionary familiarity) were 2 for Group E-P and 1 for Group P-E. It can be noted that the distribution of e-dictionary familiarity responses was bimodal: the largest number (37) was observed in Category 1 ("Never use one"), but the second largest (15) fell in Category 5 ("Several times a week"). This indicated that a substantial number of participants were already familiar with e-dictionaries while the majority of the rest had never touched one.

Table 5. Distribution of responses to Q1 (p-dictionary familiarity)

Q1	6	5	4	3	2	1	N	Median
Group E-P	4	14	6	10	4	0	38	4
Group P-E	2	8	6	14	7	2	39	3
Total	6	22	12	24	11	2	77	4

Table 6. Distribution of responses to Q2 (E-dictionary familiarity)

Q2	6	5	4	3	2	1	N	Median
Group E-P	0	9	1	6	5	17	38	2
Group P-E	2	6	0	6	5	20	39	1
Total	2	15	1	12	10	37	77	2

Since no assumptions were made concerning the underlying distributions of these data sets, the Mann-Whitney tests were conducted to examine the differences between these median pairs. The Mann-Whitney *U*s were 530.0 for Question 1 ($p < .05$) and 694.5 for Questions 2 (n.s.). Hence, although there was no indication that the two groups differed in e-dictionary familiarity, there was some evidence that Group E-P consisted of students who (had) used p-dictionaries more frequently than those in Group P-E. Rank order correlation (Spearman's *rho*) between the two variables was -.163 (n.s.). There was no indication that p-dictionary familiarity and e-dictionary familiarity were correlated in any meaningful way.

To summarize, Groups E-P and P-E were comparable in terms of overall English proficiency

and e-dictionary familiarity, though Group E-P were significantly more familiar with p-dictionaries than Group P-E were.

3.2. Speed Test Results

3.2.1. Part 1: Simple Word Search

Descriptive Statistics

Descriptive statistics for Part 1 are summarized in Table 7, in such a way that the cells involving the same dictionary type appear in the same column.

Table 7. Part 1 score means and SDs (possible max = 10)

	E-Dictionary	P-Dictionary	Combined
Form A	7.92 (1.55) Group E-P (n = 38)	5.18 (1.07) Group P-E (n = 39)	6.53 (1.91) N = 77
Form B	7.67 (1.74) Group P-E (n = 39)	5.61 (1.28) Group E-P (n = 38)	6.65 (1.84) N = 77
Combined	7.79 (1.64) N = 77	5.39 (1.19) N = 77	

In the first phase using Form A, the mean obtained using e-dictionaries was 7.90 words (out of 10) while that achieved with p-dictionaries was 5.18. Using Form B, the mean for e-dictionaries was 7.67 and that for p-dictionaries was 5.55. In the bottom row, the e-dictionary mean and the p-dictionary mean, each combining Forms A and B, are shown. In the right-most column, the Form-A mean and the Form-B mean, each combining e- and p-dictionaries, are shown. (It should be pointed out that these e-dictionary means may be underestimating e-dictionaries true potential due to ceiling effect: nine participants in Group E-P and eight in Group P-E scored 10, the possible maximum. Had more items been provided, these 17 students' scores may have been higher. Also note in passing that coefficients alpha were not computed because they are not suitable statistics when the test is speeded.)

Visual inspection of these means appears to indicate that differences between forms are not so clear but those between e- and p-dictionaries are pronounced. However, it should be reminded that the combined mean for e-dictionary, as well as that for p-dictionary, is obtained by averaging across data produced by two different groups taking two different forms. Similarly, the combined mean for Form A, as well as for Form B, is based on data produced by two different groups using two different types of dictionaries. Therefore, we cannot be certain that our impression is indeed correct until variables other than dictionary type are controlled for, which is possible in multiple regression analyses.

Preliminary Multiple Regression

To examine the effect of dictionary type when potential differences between test forms and subject groups are controlled for, multiple regression analyses were attempted with Part 1 score as the predicted variable. Predictor variables entered into the model are shown in Table 8. Categorical variables coded as 1 or 0 are known as *dummy variables*. This coding makes it possible to include nominal variables such as dictionary type or test form as regression terms (see Nunnally & Bernstein, 1994; Crown,1998).

Table 8. Independent variables tried for the preliminary model (Part 1)

- DICT: dictionary type (1 if electronic; 0 if printed)
- FORM: test form (1 if Form A; 0 if Form B)
- GROUP: subject group (1 if Group P-E; 0 if Group E-P)
- DICT*FORM; dictionary type x test form interaction (1 if electronic and Form A; 0 if otherwise)
- DICT*GROUP: dictionary type x subject group interaction (1 if electronic and Group P-E; 0 if otherwise)
- FORM*GOUP: test form x subject group interaction (1 if Form A and Group P-E; 0 if otherwise)

For variable selection, the stepwise method was used. It begins with entering into the model the variable with the strongest simple correlation with the predicted variable, and then goes on to enter the other variables one by one in the order of partial correlation strength, each time testing each variable for removal from the model. The resultant model was the following:

$$Part1\ Score = 2.403\ DICT + 5.390$$

Adjusted R-squared was .411 and the F-value was 107.929 ($p < .01$). FORM, GOUP, or the interaction terms were not significant predictors when DICT was partialed out. Namely, the only significant predictor of Part 1 score was the dictionary type. When using e-dictionaries (DICT = 1), the model reduces to: *Part 1 Score = 7.793*. When using p-dictionaries (DICT = 0), the model reduces to: *Part 1 Score = 5.390*. Hence, this model stipulates that, other things being equal, using e-dictionaries will result in Part 1 Scores that is 2.403 points higher than using p-dictionaries.

Including Background Variables

The preliminary analysis above indicated that the difference in dictionary type indeed resulted

in significantly different Part 1 scores. The next question was whether this would still hold when background variables, i.e., c-test score, p-dictionary familiarity, e-dictionary familiarity were controlled. As a next step, therefore, the variables in Table 9 were tried as predictors.

The stepwise selection procedure removed all but DICT and DICT*E_FAM. Adjusted R-squared was .483 and the *F*-value was 72.505 ($p < .01$). Regression coefficients are shown in Table 10.

Table 9. Independent variables tried for the final model (Part 1)

- DICT: dictionary type (1 if electronic; 0 if printed)
- C-TEST: C-test number correct
- P_FAM: p-dictionary familiarity rating (6-1).
- E_FAM: e-dictionary familiarity rating (6-1)
- DICT*C_TEST: interaction between dictionary type and c-test score
- DICT*P_FAM: interaction between dictionary type and p-dictionary familiarity
- DICT*E_FAM: interaction between dictionary type and e-dictionary familiarity

Table 10. Regression coefficients in the final model (Part 1)

	Unstandardized		Standardized		
	B	Std. Error	Beta	t	P
(Constant)	5.390	0.153		35.171	0.000
DIC_X_EF	0.438	0.093	0.392	4.701	0.000
DICT	1.356	0.311	0.364	4.365	0.000

The final model was:

$$PART1 = 5.390 + 0.438\ DICT*E\_FAM + 1.356\ DICT$$

For e-dictionary (DICT = 1), the model reduces to:

$$PART1 = 0.438\ E\_FAM + 6.755$$

For p-dictionary (DICT = 0), the model reduces to:

$$PART1 = 5.390$$

The two regression lines, one for e-dictionaries and the other for p-dictionaries, are graphically shown in Figure 1.

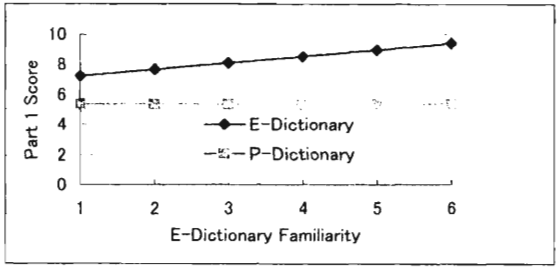

Figure 1. Regression lines predicting Part 1 score from e-dictionary familiarity

Cross Validation

To cross-validate the obtained model, (1) the participant group was randomly divided into two, (2) separate regressions were run on each subgroup, using DICT and DICT*E_FAM as predictors, and (3) obtained regressions were compared to see if they could be considered equivalent. The obtained statistics for the two subgroups are shown, alongside those for the whole group, in Table 11.

From visual inspection of the coefficients and *t*-statistics, it appears that the models for two subgroups are quite similar to each other and to that for the whole sample. A formal test of the equality of these two regression equations was conducted by the method based on sum of squared errors (Crown, 1998, pp. 46-7). The *F*-test yielded an *F* of 0.431 (*df* = 3,148; n.s.). Thus, the null hypothesis that there is no difference between these two regressions was not rejected. Hence, the model obtained from the whole sample was shown to be a stable one.

Table 11. Regression coefficients obtained for two subgroups (Part 1)

	Whole Group		Subgroup A		Subgroup B	
Variable	Coefficient	t	Coefficient	t	Coefficient	t
DICT	1.356	4.365	1.054	2.297	1.724	4.122
DICT*E_FAM	0.438	4.701	0.552	4.050	0.276	2.244
Constant	5.390	35.171	5.282	23.341	5.500	26.529
Adjusted R^2	0.483		0.476		0.482	
F	72.505		35.962		36.382	
Sig.	p = .000		p = .000		p = .000	

All in all, then, it was confirmed that (1) using e-dictionaries indeed resulted in significantly higher Part 1 scores, and (2) the effect was multiplied by users' degree of familiarity with e-dictionaries. More specifically, even with students who have never used e-dictionaries (E_FAM = 1), the predicted score using e-dictionaries (6.755 + 0.438*1 = 7.094) is significantly higher than that using p-dictionaries (5.39), and one degree increase in e-dictionary familiarity is predicted to

add 0.438 points increase in Part 1 score, resulting in a score of 9.383 (= 0.438*6 + 6.755), when E_FAM is the highest (= 6). English proficiency as measured by the c-test or familiarity with p-dictionaries did not affect this relationship.

3.2.2. Part 2: Example Search

Descriptive Statistics

Descriptive statistics for Part 2 scores (the number of examples located) are summarized in Table 12. For each cell, the possible maximum mean is 5.00. The Form A mean (= 2.26) was approximately 0.5 points lower than the Form B mean (=2.77), and the e-dictionary mean (= 2.43) was somewhat lower than the p-dictionary mean (= 2.60). Again we are not sure whether these differences derive from test form, dictionary type, subject group, or some interaction, until we sort them out by multiple regressions.

Table 12. Part 2 score means and SDs (possible max = 5)

	E-Dictionary	P-Dictionary	Combined
Form A	2.29 (1.61) Group E-P (n = 38)	2.23 (1.55) Group P-E (n = 39)	2.26 (1.57) N = 77
Form B	2.56 (1.48) Group P-E (n = 39)	2.97 (1.20) Group E-P (n = 38)	2.77 (1.36) N = 77
Combined	2.43 (1.54) N = 77	2.60 (1.43) N = 77	

Multiple Regressions

As in the case of Part 1 score, we first entered DICT, FORM, GROUP, DICT*FORM, DICT*GROUP, and GROUP*FORM to predict Part 2 score, using the stepwise selection method. This time, all variables were removed but FORM, meaning that dictionary type, group difference, or the interactions did not contribute to the prediction. For all the efforts to make them equivalent, Part 2s of Form A and Form B were not exactly comparable, the former being significantly more difficult. Concerning our main interest, dictionary type difference, there was no evidence that use of e-dictionaries or p-dictionaries resulted in significantly different Part 2 scores.

As a next step again, background variables (C-TEST, P-FAM, E-FAM, DICT*C-TEST, DICT*P-FAM, DICT*E-FAM) were entered together with FORM to go through stepwise variable selection. Only E-FAM and FORM were retained in the final model. Adjusted R-squared was minimal (= .048), though the model was significant (F = 4.873, p < .01). The regression coefficients are shown in Table 13. It can be seen that E_FAM (t = 2.242) contributed somewhat more than FORM (t = -2.172).

Table 13. Regression coefficients in the final model (Part 2)

	Unstandardized		Standardized		
	B	Std. Error	Beta	t	P
(定数)	2.386	0.236		10.096	0.000
E_FAM	0.159	0.071	0.177	2.242	0.026
FORM	-0.506	0.233	-0.171	-2.172	0.031

Hence, the final model for Part 2 score was the following:

Part 2 Score = 0.159 E_FAM - 0.506 FORM + 2.386

This model was rather counterintuitive. It stipulates that familiarity with e-dictionaries is positively correlated with efficiency of example search, *even when using p-dictionaries*. Cross validation was conducted in the same manner as Part 1 score: using FORM and E_FAM as predictors, separate regressions were run on two randomly divided subgroups, to see if similar models would emerge. The resultant coefficients are shown in Table 14.

Table 14. Regression coefficients obtained for two subgroups (Part 2)

	Whole Group		Subgroup A		Subgroup B	
Variable	Coefficient	*t*	Coefficient	*t*	Coefficient	*t*
E_FAM	0.159	2.242	0.241	2.460	0.030	0.295
FORM	-0.509	-2.172	-0.388	-1.181	-0.586	-1.781
Constant	2.386	10.096	2.271	6.766	2.545	7.640
Adjusted R^2	0.048		0.066		0.017	
F	4.873		3.668		1.653	
Sig.	$p = .009$		$p = .030$		$p = .198$ (n.s.)	

For subgroup A, the coefficient for FORM was not significant (t = -1.181, n.s.), though the model overall was (p = .030). For subgroup B, neither the coefficient for E_FAM (t = 0.295) nor that for FORM (t = -1.781) was significant. Hence, the model obtained for the whole sample was not validated for either subgroup. Therefore, that E_FAM emerged as a significant predictor for the whole sample had better be considered an artifact of chance. The only consistent finding was that for the whole group, or either of the subgroups, dictionary type was not a significant predictor, meaning that it did not make any difference on Part 2 scores whether e- or p-dictionaries were used.

3.2.3. Learners' Impressions

The responses to impression questions are summarized in Tables 15-20. To recapitulate, the rating "5" implies the highest degree of the characteristic in question.

Table 15. Responses to impression question 3: Ease of locating specified entries

	5	4	3	2	1	Median
P-Dictionary	0	15	40	21	1	3
E-Dictionary	36	29	7	5	0	4

Table 16. Responses to impression question 4: Ease of locating specified examples

	5	4	3	2	1	Median
P-Dictionary	3	11	22	28	13	2
E-Dictionary	6	9	16	34	12	2

Table 17. Responses to impression question 5: Eye-friendliness of located entries

	5	4	3	2	1	Median
P-Dictionary	3	15	27	28	4	3
E-Dictionary	8	31	24	12	2	4

Table 18. Responses to impression question 6: Willingness to use

	5	4	3	2	1	Median
P-Dictionary	1	9	25	35	7	2
E-Dictionary	26	37	10	4	0	4

Table 19. Responses to impression question 7: Usefulness for learning English

	5	4	3	2	1	Median
P-Dictionary	7	32	32	6	0	4
E-Dictionary	33	33	8	3	0	4

Table 20. Responses to impression question 8: Overall impression

	5	4	3	2	1	Median
P-Dictionary	3	18	47	8	1	3
E-Dictionary	23	40	11	3	0	4

Wilcoxon signed-rank tests revealed that for all the response-group pairs but Question 4, the median for e-dictionary was significantly ($p < .01$) higher. That is to say, participants felt that e-dictionaries (1) made it easier to access word meanings, (2) displayed information in a more eye-friendly manner, (3) made them more willing to use dictionaries, (4) were more useful for learning English in general, and (5) were better overall, than p-dictionaries. With regard to locating examples, participants' median responses were "2" ("Rather difficult") for both types of dictionaries.

3.2.4. Familiarity and Impression

Finally, whether there was a meaningful relationship between the pre-study familiarity with p- or e-dictionaries and the post-experiment impression of each dictionary type was explored.

Cross-tabulation of familiarity with e-dictionaries (E_FAM) by overall impression of e-dictionaries (E_IMP) did not visually imply any relationship between the two ratings (Table 21). In fact, neither Kendall's *tau*-b (.146, n.s.) nor Spearman's *rho* (.165, n.s.) reached significance.

The other three familiarity-impression cross-tabulations (P_FAM by E_FAM, P_FAM by P_IMP, and E_FAM by P_IMP) were all tried, but none of them revealed any significant relationships. Therefore, participants' impressions of e- and p-dictionaries were not related to their pre-experiment familiarity with each dictionary type. Hence the tendency for users' familiarity with each dictionary type to affect their impressions (Koyama & Takeuchi's 2001a) was not identified in the present study.

Table 21. Cross-tabulation: Familiarity by impression [E-dictionaries]

	E_IMP					
E_FAM	5 (very good)	4 (rather good)	3 (mixed)	2 (rather bad)	1 (very bad)	Sum
6 (use every day)	1	1	0	0	0	2
5 (2-3 times a week)	6	7	2	0	0	15
4 (once a week)	1	0	0	0	0	1
3 (occasionally use)	4	7	1	0	0	12
2 (rarely use)	1	6	3	0	0	10
1 (have never used)	10	19	5	3	0	37
Sum	23	40	11	3	0	77

4. Discussion and Conclusion

Before discussing the results, limitations of this study should be pointed out. Part 1s in both forms were deliberately written in such a way that the length of words would gradually increase from Item 1 to Item 10. In retrospect, this was not the right thing to do. Placing easier items before more difficult ones may be a common and reasonable practice for a power test, but not for a speed test like the one for the study. Secondly, the Part 2 task was only an indirect measure of real-life example search efficiency. Cross-validation using more authentic tasks are clearly in order.

Notwithstanding these limitations, the study is believed to have placed pocket e-dictionaries in a better light. It has done so not as much by discovering previously unknown phenomena as empirically confirming what has been informally believed to be the case.

Part 1 results coincided with the finding by Tsuchimochi (2002) that pocket e-dictionaries enable users to access words and identify their meanings significantly more quickly than p-dictionaries. The present study's unique contribution to the field is that it has specified a regression model incorporating familiarity factor. Word search efficiency was found to be a

function of dictionary type and familiarity with e-dictionaries. Even when learners have never used e-dictionaries, they can still locate words on them significantly more quickly than on p-dictionaries, and as they get more accustomed to using them, the electronic-printed gap widens proportionately. It is noteworthy that neither general English proficiency nor familiarity with p-dictionaries was found to be a significant factor once the two variables above were entered into the model.

It may seem counterintuitive that familiarity with p-dictionaries was not identified as a significant factor, even for predicting word-search efficiency on p-dictionaries. This was probably because the participants of this study were university students, most of whom presumably had already acquired basic p-dictionary-using skills even when they "rarely used" p-dictionaries at the time of the experiment. Had participants been those with much shorter learning experience, the results might have been different.

Another important finding was that examples were no more difficult to locate in e-dictionaries than in p-dictionaries. This was confirmed by empirical data as well as by subjective responses to the impression survey: electronic-printed difference was not significant in Part 2 scores or in ratings to Question 4. The result constitutes counter-evidence against the preconception that examples in e-dictionaries are more difficult to explore due to hierarchical data structure.

The third, and potentially the most important, finding was that learners are markedly less reluctant to use dictionaries when they are electronic. In other words, e-dictionaries' efficient access appear to lower the "consultation trigger point" (Aust et al., 1993, p. 70). The electronic-printed gap in word search efficiency observed in this study as well as in Ttsuchimochi (2002) is one obtained in, as it were, a *forced* search situation. When compelled to do so, users are able to locate, on e-dictionaries, 1.5 - 2 times more words than on p-dictionaries. In real life, however, there will be no one to force them to use dictionaries. Therefore, there are likely to be numerous situations where learners will look up words *if and only if* the dictionaries at hand are electronic. Indeed there is a report that in reading a dictionary-incorporated hypertext passage, learners consulted the dictionary twice more frequently than when reading the same material using p-dictionaries (Aust et al., 1993). In the case of pocket e-dictionaries as well, the electronic-printed gap in the number of words learners *will* look up should be much, much larger.

Even if words looked up on e-dictionaries are *felt* to be retained somewhat less well (Koyama & Takeuchi 2002a), is it not reasonable to assume that the much larger look-up frequency resulting from the lowered "consultation trigger point" outweigh the alleged disadvantage?

Notes

This article is an elaborated version of the paper presented by the author at the 42nd Annual Conference of LET, held at Otsuma Women's University, on August 2nd, 2002. The author thanks CASIO Computer

Co., Ltd. for lending the e-dictionaries for this study and Prof. Amma Kazuo for letting him use the c-test.

References

Aust, R., Kelly, M. J., & Roby, W. (1993). The use of hyper-reference and conventional dictionaries. *Educational Technology, Research, and Development, 41*(4), 63-73.

Barnett, M. A. (1989). *More than meets the eye: Foreign language reading: Theory and practice.* Englewood Cliffs, NJ: Prentice-Hall.

Bhatia, A. (1991). Kanji retrieval by recursive of location of elements using HyperCard. *CALICO Journal, 9*(2), 4-25.

Crown, W. H. (1998). *Statistical models for the social and behavioral sciences.* Westport, CT: Praeger.

Keller, H. H. (1987). Pedagogical wishes for a machine dictionary: An example from Russian. *The Modern Language Journal, 71*(1), 12-17.

Knight, S. (1994). Dictionary use while reading: the effects on comprehension and vocabulary acquisition for students of different verbal level. *The Modern Language Journal, 78*(3), 285-299.

Koga, T. (1995). The effectiveness of using an electronic dictionary in second language reading. *Bulletin of the Faculty of Education, Hiroshima University, Part II (Arts and Science Education), 44,* 239-244.

Koyama, T. & Takeuchi, O. (2001a). How students differ in using dictionaries: printed dictionaries vs. electronic dictionaries. *Proceedings of the 41st Annual Conference of Japan Association for Language Education and Technology*, 196-199.

Koyama, T. & Takeuchi, O. (2001b). Qualitative comparison of searching behavior: printed dictionaries vs. electronic dictionaries. *Proceedings of the 17th Annual Conference of Japan Society for Educational Technology, 2001*, 477-178.

Koyama, T. & Takeuchi, O. (2002a). Printed dictionaries vs. electronic dictionaries: how EFL students make use of information from each type. *Proceedings of the 42nd Annual Conference of Japan Association for Language Education and Technology*, 108-109.

Koyama, T. & Takeuchi, O (2002b). Printed dictionaries vs. electronic dictionaries: influence of interface design on the searching behavior by Japanese EFL learners. *Proceedings of the 18th Annual Conference of Japan Society for Educational Technology*, 811-812.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: CUP.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory.* New York: McGraw-Hill.

Takada, T. (2002, November). *Koukou ni okeru densi jisho no katuyou ni tuite* [On using electronic dictionaries at high school level]. Paper presented at the Seminar on Using Electronic Dictionaries in High School Classes, Kyoto, Japan.

Tsuchimochi, K. (2002) *On the use of electric dictionaries and paper dictionaries.* Manuscript in preparation, Hiroshima University.

Yanase, Y. (2002, May). *Literacy being lost.* Paper presented at the Seminar on Using Electronic Dictionaries in High School Classes, Hiroshima, Japan.