Does dictionary form really matter? Anna Dziemianko

Adam Mickiewicz University danna@ifa.amu.edu.pl

1. Introduction

A number of experimental studies have been devoted to the relative usefulness of paper and electronic dictionaries (e.g., Chen 2010, Dziemianko 2010, Kobayashi 2007, 2008, Koyama and Takeuchi 2003, 2004, 2007, Nesi 2000, Osaki et al. 2003, Shizuka 2003, Xu 2010). Nonetheless, neither these studies nor the many others which investigate the potential of the electronic medium give a clear indication of whether paper or electronic dictionaries are more useful in specific language activities. Unfortunately, comparability issues, related, among other things, to samples, dictionaries and tasks, pose a serious obstacle to any systematic comparative analysis necessary to generalize on the basis of the pertinent literature. In fact, it is virtually impossible to draw any exact parallels between studies concerned with similar issues, let alone the results they yield (Dziemianko: In press). It appears, then, that replicating research might be helpful inasmuch as it could enable careful comparison and make it possible to see whether the conclusions drawn from the reproduced studies are context-bound (e.g., sample-specific, dictionary-specific or task-specific) or not.

2. Methods

2.1. The aim of the study

The research presented below is a replication of the study conducted by Dziemianko (2010), which led to the conclusion that e-COBUILD6, available online by a pin code, was much more useful in reception and production than COBUILD6 in paper form. The electronic version of the dictionary also proved more helpful in learning meaning and collocations. The aim of the replication was to see whether similar results can be obtained for LDOCE5.

2.2. Materials

The materials from the original research were used, except for the dictionary. Thus, the same pretest, test proper, retention test and questionnaire as in Dziemianko (2010) were employed. In the test, there were both receptive and productive tasks. In the receptive part of the test, the subjects were requested to explain the meaning of the following nine nouns and phrases: backgammon, booby prize, clampdown, collateral damage, down under, dream ticket, flapjack, onus and outcrop. They could use either English or Polish. In the encoding task, prepositions removed from as many collocations incorporated into sentences had to be supplied: on the blink, in cahoots, up the creek, at gunpoint, wreak havoc on, in the offing, in the pipeline, under sedation, on the trot. To perform the tasks, each subject consulted LDOCE5, either in paper or electronic form (the free online version). The test proper was the same as the pretest and the retention test, with that in the latter, the order of the target items was changed to prevent the learning effect. The subjects' familiarity with dictionary formats was investigated by means of a survey. The participants were asked whether they were used to consulting monolingual learners' dictionaries (henceforth MLDs), and, if yes, whether they referred to paper or electronic versions thereof, or both.

2.3. Subjects and procedures

87 Polish students of English participated in the research (B2-C1 according to the CEFR, like in Dziemianko (2010)). 42 students consulted LDOCE5 in book form, and the other 45 used e-LDOCE5. The experiment started with the pretest, where the receptive and productive tasks had to be performed without access to any dictionary. The cases where the

participants knew correct answers were eliminated from further analysis. The subjects were also requested to fill in the accompanying questionnaire. Once the pretest with the questionnaire had been collected, tests sheets were distributed. The students were instructed to do the tasks again using the dictionary version to which they had been assigned. The experiment was preceded by a short orientation session, in which the participants were familiarized with LDOCE5 online and on paper. After two weeks, retention was checked in an unexpected recall test, in which no dictionaries were allowed.

3. Results

3.1. The questionnaire

The information obtained from the questionnaire makes it clear that all the subjects consulted MLDs as a matter of routine. In both experimental conditions (i.e., in the groups working with paper and electronic LDOCE5, henceforth PD group and ED group, respectively), comparable proportions of students were familiar with paper and electronic formats (PD group: paper dictionary users 76.2 percent, electronic dictionary users 69.0 percent, p=0.76; ED group: paper dictionary users 57.8 percent, electronic dictionary users 60.0 percent, p=0.85; Z test for dependent observations, non-significant, alpha-level=0.05). Thus, familiarity with paper and electronic MLDs could not have affected task accomplishment in the experiment.

3.2. Test proper and retention test results

A summary of the subjects' performance in the receptive and productive tasks in both tests is presented graphically in Figure 1. Table 1 gives the results of the repeated measures ANOVAs conducted for both tests, with the task as the within-group factor.

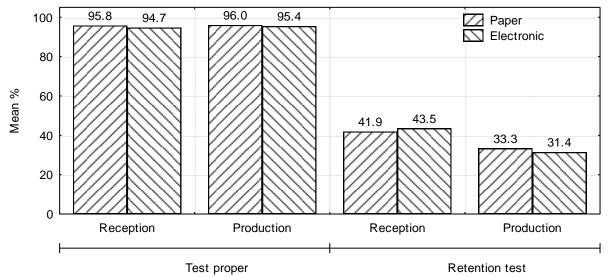


Figure 1. Test proper and retention test results (LDOCE5)

Table 1. Test proper and retention test results (LDOCE5)

	7	Test	Retention		
	F	р	F	p	
Form	0.2	0.664	0.0	0.983	
TASK	0.0	0.868	2.0	0.176	
TASK*Form	0.0	0.919	0.1	0.815	

First, the scores achieved in reception and production in the test proper were comparable among the users of paper and electronic LDOCE5 (alpha-level=0.05). The medium had no statistically significant effect on passive and active recall, either; the subjects

using e-LDOCE5 were as successful in recalling meaning and collocations as those who consulted p-LDOCE5. Second, the task had no considerable influence on the obtained results; in each test, the subjects' scores were comparable in both tasks. Finally, the interaction between the task and dictionary form did not produce any statistically significant effect on the subjects' performance at any stage of the experiment.

4. COBUILD6 vs. LDOCE5

4.1. Test proper

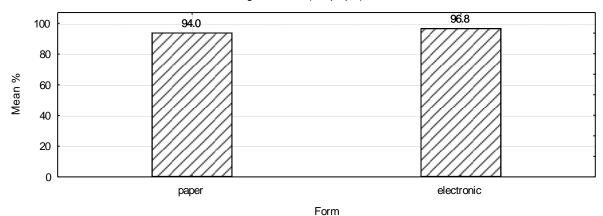
Obviously, the obtained results diverge from those presented by Dziemianko (2010).¹ To find out more about the differences, Table 2 gives results of the repeated measures ANOVA for test proper scores in both experiments, with dictionary as a between-group factor.

Table 2. ANOVA summary table: COBUILD6 vs. LDOCE5 (test proper)

	F	p
Dictionary	0.0	0.884
Form	7.0	0.013*
Dictionary*Form	11.7	0.002**
TASK	0.0	0.871
TASK*Dictionary	0.0	0.887
TASK*Form	0.0	0.997
TASK*Dictionary*Form	0.0	0.856

The ANOVA reveals that differences between test proper results can be put down to the medium as well as its interaction with the dictionaries used in the two studies. Figure 2 illustrates graphically the role of the medium.

Figure 2. Form (test proper)



The analysis of the main effect suggests that the electronic medium considerably facilitated dealing with the test. Although the mean score obtained with the help of the electronic dictionaries was only three percent higher than the average result of paper dictionary consultation, the difference was statistically significant, as shown in Table 2.

To see how the influence of the medium depended on the dictionaries used, the results of the Tukey HSD test are collated in Table 3. The mean proportions compared are illustrated in Figure 3.

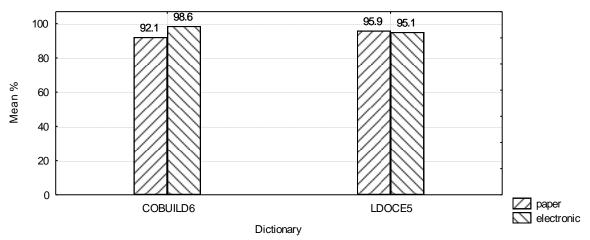
Table 3. The Tukey HSD: Dictionary x Form (test proper)

	Mean %	Dictionary	Form	1	2	3	4
1	92,1	COBUILD6	paper		0.001**	0.076	0.218
2	98,6	COBUILD6	electronic	0.001**		0.307	0.116
3	95,9	LDOCE5	paper	0.076	0.307		0.947
4	95,1	LDOCE5	electronic	0.218	0.116	0.947	

¹ See section 2.1.

-

Figure 3. Dictionary*Form (test proper)



Clearly, the medium played a role only in the case of COBUILD6; the scores obtained with the help of e-COBUILD6 were seven percent higher and significantly better than those achieved by p-COBUILD6 users. The same does not hold for the paper and electronic versions of LDOCE5, which did not seriously affect the subjects' performance in the test.

4.2. Retention test

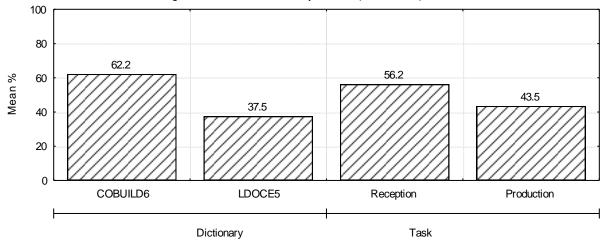
The results of the repeated measures ANOVA for retention test scores in both experimental sessions are given in Table 4.

Table 4. ANOVA summary table: COBUILD6 vs. LDOCE5 (retention test)

	F	p
Dictionary	25.0	0.000**
Form	2.7	0.112
Dictionary*Form	2.8	0.103
TASK	11.7	0.002**
TASK*Dictionary	0.4	0.545
TASK*Form	0.0	0.972
TASK*Dictionary*Form	0.3	0.616

As can be seen, retention differences between the two experiments were caused by the dictionaries and the task. Figure 4 supplies further details on the role of these factors.

Figure 4. Main effects: Dictionary and Task (retention test)

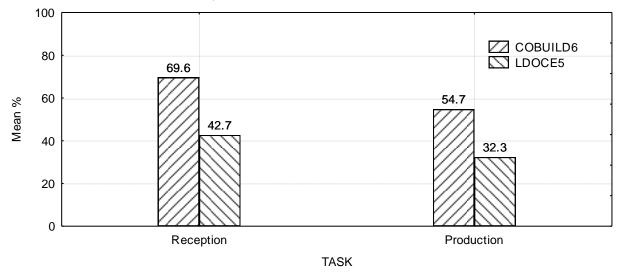


Overall, retention was 66 percent better after reference to COBUILD6 than LDOCE5. Besides, passive recall proved to be generally much easier than active recall; about one third more subjects remembered meaning than collocations. Since, as shown in Table 4, the interaction between the task and the dictionary is not statistically significant (p=0.545, alpha level=0.05), but the main effects produced by the task and the dictionary are, it may be expected that the scores in the productive and receptive tasks differ between COBUILD6 and LDOCE5, with that the inter-dictionary differences, though significant, are comparable. To get more information about the interaction, results of the Tukey HSD test are presented in Table 5 and illustrated graphically in Figure 5.

Table 5. The Tukey HSD: Task*Dictionary (retention test)

	Mean %	Dictionary	TASK	1	2	3	4
1	69,6	COBUILD6	Reception		0.036*	0.003**	0.000**
2	54,7	COBUILD6	Production	0.036*		0.218	0.015*
3	42,7	LDOCE5	Reception	0.003**	0.218		0.216
4	32,3	LDOCE5	Production	0.000**	0.015*	0.216	

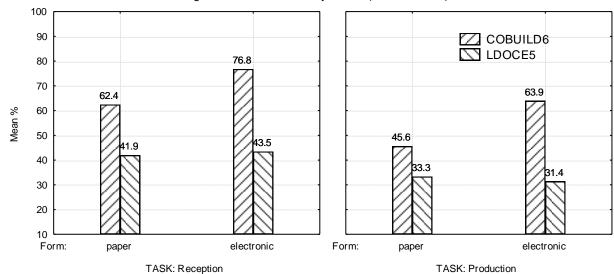
Figure 5. TASK*Dictionary (retention test)



As can be seen, in the retention test, the results obtained after reference to COBUILD6 were always better than those which followed from the use of LDOCE5. Indeed, as expected, the differences between the two dictionary conditions where in each task comparably big and statistically significant (63 percent in passive recall, p=0.003; 69 percent in active recall, p=0.015, alpha level=0.05). Thus, the dictionary did affect the subjects' results in the retention test, with that the effect was comparable for reception and production.

In view of the fact that in the study by Dziemianko (2010), where COBUILD6 was used, the role of dictionary form (paper vs. electronic) was highly significant, but it proved to have no statistically significant bearing on retention in the case of LDOCE5 (Table 1, p=0.983, alpha level=0.05), it might be interesting to delve deeper into the data and see how dictionary form affected the interaction between task and dictionary, the lack of statistical significance notwithstanding (Table 4, p=0.616, alpha level=0.05). Figure 6 illustrates the relevant data.

Figure 6. TASK*Dictionary*Form (retention test)



When the subjects used the online versions of the dictionaries, those who referred to e-COBUILD6 remembered considerably more words (77 percent) and collocations (104 percent) than those who consulted e-LDOCE5 (Tukey HSD; reception: p=0.036, production: p=0.044, alpha level=0.05). However, differences either in passive or active recall (49 percent and 37 percent, respectively), although still in favor of COBUILD6, were not statistically significant when the paper versions of the dictionaries were used (Tukey HSD; reception: p=0.449, production p=0.913, alpha level=0.05).

5. Discussion

The results obtained by Dziemianko (2010) were not confirmed for LDOCE5. For one thing, the medium had no statistically significant bearing on reception or production. For another, it did not affect the retention of meaning and collocations, either. Such findings stand in stark contrast to those presented in Dziemianko (2010), where the electronic version of COBUILD6 proved much more conducive to decoding and encoding, and largely facilitated passive and active recall. On the surface, COBUILD6 seems overall more helpful than LDOCE5 in the retention of both meaning and collocations. Yet, a more detailed, interdictionary comparison reveals significant differences in retention results only for the electronic versions of the dictionaries. Thus, it is necessary to reflect on two aspects:

- 1. why e-LDOCE5 did not prove to be more useful in reception, production and recall than LDOCE5 on paper (in contrast to what holds true for COBUILD6),
- 2. what makes e-COBUILD6 more helpful in retention than e-LDOCE5.

The screenshots presented in the appendix might help address both questions. Obviously, e-COBUILD6 is much clearer and more neatly organized than e-LDOCE5. In the latter, there is an excess of noise in the form of colorful widgets as well as banners employing video and animation, embedded into the web page above the headword and below the definition. Obviously, they deflect users from the dictionary itself not only by animation, but also vibrant, garish colors. By the same token, they make dictionary information much less salient on the glutted website. Tower adverts, less gaudy and usually related to the page being viewed, but rotating with every new search like the aforementioned banners, are displayed in the right-hand margin of the web page. They no doubt further contribute to dwarfing lexicographic data. By contrast, no advertisements are allowed on the relatively plain and simple website of e-COBUILD6. As a result, dictionary information is much more conspicuous there. The e-COBUILD6 search window is followed by the entry for the word

looked up, and on the right there are only a few buttons which users might click if they want to expand their vocabulary, customize the dictionary or get help. Presumably, the e-LDOCE5 disruptive and overwhelming noise, which makes the important information difficult to discern and concentrate on, could have been a factor accounting for the better results obtained with the help of e-COBUILD6 than e-LDOCE5.

The abundance of unsolicited information in loud colors and different shapes might also be a reason why e-LDOCE5 proves to be no more useful than LDOCE5 on paper. As suggested by Dziemianko (in print), the largely automatic stages of tracking down relevant entries in paper dictionaries, which usually require mechanical effort rather than cognitive exertion, correspond to what Craik and Lockhart (1972: 673) consider an initial phase of stimuli examination preceding cognitive analysis. Nonetheless, as they point out, this early juncture hardly contributes to strengthening the memory trace, which is dependent not so much on mechanical engagement as on semantic involvement. However, since the preliminary examination of stimuli involves "analysis of such physical or sensory features as lines, angles, brightness, pitch, and loudness" (Craik and Lockhart 1972: 673), it no doubt applies also to electronic dictionaries. Obviously, such stimuli in the form of noise were in profusion in e-LDOCE5. The obtained results tentatively suggest that they might have been no less absorbing than wading through alphabetically arranged entries on paper. In fact, they might have largely counterbalanced the beneficial effect of rapid outer access in electronic dictionaries, which leads users directly to the desired headword without the need to go through running heads (Bergenholtz and Gouws 2007: 244). Apparently, neither the paper dictionary search nor the obtrusive noise in the electronic dictionary induced semantic involvement which could inform retention.

Apart from the disruptive influence of noise, entry navigation in e-LDOCE5 might account for the fact that, in contrast to Dziemianko's (2010) findings for COBUILD6, this online dictionary does not prove to be any better for language reception in the test proper than LDOCE5 on paper. Unlike e-COBUILD6, e-LDOCE5 gives no hits for the phrases *down under* and *collateral damage*. Additional effort is required from users to access them at *down* ¹28 and *collateral* ²1, respectively, going through clickable menus first. Importantly, no menus are given for the corresponding headwords in LDOCE5 on paper. Item analysis reveals above-average score differences for both phrases between paper and online LDOCE5 users; in the PD group, the score was 13 percent better for *down under* and two percent better for *collateral damage*, the mean difference between the two dictionary conditions being 1.6 percent. Although the differences were not statistically significant (p=0.349 and p=0.594, respectively, Z test for independent samples, alpha level=0.05), they were above average, which suggests that the additional effort invested into finding the phrases in e-LDOCE5 because of the menus might have been detrimental to reception.

However, e-LDOCE5 menus do not seem to have affected performance in the production task in the test proper, where they guided the subjects to *in the offing*, *on the blink* and *on the trot*. To retrieve these collocations, e-LDOCE5 users had to identify the right match in the menu, click on it and, apart from *in the offing*, find the right sense in the relevant entry ($trot^2$ 2 and $blink^2$ 1). Even though only the phrase *in the offing* is explained in the entry for *offing*, to access the collocation users first had to face a menu with as many as 90 matches. In the paper version of LDOCE5, no menus serve to navigate the corresponding entries. Item analysis reveals that for *in the offing*, success rates were the same in the groups dealing with LDOCE5 online and on paper, and for *on the blink*, they were one percent higher in the PD group (p=0.847, Z test for independent samples, alpha level=0.05, non-significant). The responses for *on the trot*, in turn, were three percent more often correct among e-LDOCE5 users (p=0.499, Z test for independent samples, alpha level=0.05, non-significant). However, in these cases, the differences not only lacked statistical significance, but they did not exceed

the average inter-dictionary difference for a production task item, which amounted to 3.3 percent.

Item analysis sheds some light on possible reasons for the overall inferior retention scores obtained with the help of e-LDOCE5 in comparison with e-COBUILD6. It seems that accessing the relevant information through clickable menus in e-LDOCE5 is detrimental to passive recall and, above all, active recall. It is worth remembering that there are no menus in e-COBUILD6, and users find the target nouns, phrases and collocations in relevant (sub)entries, all of which are displayed at a time.² In the case of passive recall, the results obtained in the e-COBUILD6 group for collateral damage and down under were 5 percent and 97 percent better, respectively, than those achieved in the e-LDOCE5 group. The menuless access paths in e-COBUILD6 might also be a reason why the aforementioned three collocations from the productive task (on the trot, on the blink and in the offing) were better retained by the dictionary users than by the e-LDOCE5 group, who had to cope with menus. The results in the e-COBUILD6 group were over five times (on the trot), two times (on the blink) and 75 percent (in the offing) better. Assuming that the effort invested into coping with the hierarchical, step-wise outer access structure in electronic dictionaries is mechanical rather than cognitive, i.e., it is not directly related to processing *relevant* semantic information, it is not surprising that it does not contribute to strengthening the memory trace, but, quite the reverse, apparently hinders retention.

Item analysis brings out another interesting observation. In the case of *backgammon*, retention was 21 percent better among e-LDOCE5 users than among those who consulted e-COBUILD6. A possible reason for this difference is a picture of the game in the former dictionary, absent from the latter, which might have had a facilitative effect on retention. Such a conclusion is in keeping with the findings by Chun and Plass (1996), who observed that static pictures enhance lexical recall. Yet, in the present investigation, consultation of specific entry components was not logged, thus the above remark is only a tentative one and should rather be treated as a hypothesis which requires verification.

It is worth pointing out that the target items were defined by means of contextual definitions in COBUILD6, and, except for *in the pipeline* (*pipeline2*) – with the help of analytical ones in LDOCE5. The same definitions feature in the paper and electronic versions of each dictionary. Yet, the defining styles cannot be held responsible for the difference in the usefulness of the online dictionaries for retention. If they had been a factor, they would have affected the subjects' retention based on paper dictionary consultation as well, which was not the case.⁴

It seems that further research into the use of electronic dictionaries should address the issue of noise on dictionary websites. Admittedly, it is advertisements that make it possible for anyone to access online dictionaries for free. Yet, it is by no means certain that a dictionary website crammed with advertisements remains as useful as one without them. The tentative conclusion following from the present investigation is that unsolicited promotional material can lose an online dictionary much of its usefulness. Only further empirical research with noise as a strictly controlled variable can verify this claim.

² For none of the target items does e-COBUILD6 list what Lew (in press) calls partial entries, where only the beginning of each entry is displayed with a link (*More*) to complete information. Thus, the dictionary did not require any additional clickwork in the experiment.

³ Unfortunately, the relatively small number of observations (N<30) for these collocations as well as *collateral damage* and *down under* among e-COBUILD6 users prevents running the Z test. It is obvious, however, that the cross-item differences in active recall are more conspicuous than those in passive recall. It might also be interesting to note that active recall scores were better by over half (*in the offing*) and four fifths (*on the trot* and *on the blink*) among e-LDOCE5 users than among p-LDOCE5 users, the average difference per item being only 13.5 percent.

⁴ See section 4.2. Figure 6.

Another area that requires investigation is the effect of the hierarchical nature of data display in electronic dictionaries on retention. The study was not designed to look into the role of clickable menus. However, the preliminary assessment of the possible influence of such menus on retention, and active recall in particular, is quite pessimistic. It might thus be instructive to investigate their effect more methodically and distinguish between passive and active recall to see whether they are indeed more detrimental in one respect than in the other. Besides, it is by no means certain whether clickable menus affect language reception and production. The divergent ancillary findings from the present study should also be verified in further research.

6. References

- Bergenholtz, H. and Gouws, R. (2007). "The access process in dictionaries for fixed expressions". *Lexicographica*. *International Annual for Lexicography* 23. 237-260.
- Chen, Y. (2010). "Dictionary use and EFL learning: A contrastive study of pocket electronic dictionaries and paper dictionaries". *International Journal of Lexicography* 23 (3). 275-306.
- Chun, D. M. and Plass, J. L. (1996). "Effects of multimedia annotations on vocabulary acquisition". *The Modern Language Journal* 80 (2). 183-198.
- Collins COBUILD Advanced Dictionary. 6th ed. Boston: Heinle Cengage Learning, Glasgow: Harper Collins Publishers. 2008. [COBUILD6].
- Collins COBUILD Advanced Dictionary. [online 6th ed.]. http://www.myCOBUILD.com. [Access date 20. Feb. 2011].
- Craik, F. I. M. and Lockhart, R. S. (1972). "Levels of processing: A framework for memory research". *Journal of Verbal Learning and Verbal Behavior* 11. 671-684.
- Dziemianko, A. (2010). "Paper or electronic? The role of dictionary form in language reception, production and the retention of meaning and collocations". *International Journal of Lexicography* 23 (3). 257-273.
- Dziemianko, A. (In press). "On the use(fulness) of paper and electronic dictionaries". In Granger, S. and Paquot, M. (eds.). *Electronic Lexicography*. Oxford University Press.
- Kobayashi, C. (2007). "Comparing electronic and printed dictionaries: Their effects on lexical processing strategy use, word retention and reading comprehension". In Bradford-Watts, K. (ed.). *JALT 2006 Conference Proceedings*. Tokyo: JALT. 657-671.
- Kobayashi, C. (2008). "The use of pocket electronic and printed dictionaries: A mixed-method study". In Bradford-Watts, K., Muller, T. and Swanson, M. (eds.). *JALT 2007 Conference Proceedings*. Tokyo: JALT. 769-783.
- Koyama, T. and Takeuchi, O. (2003). "Printed dictionaries vs. electronic dictionaries: A pilot study on how Japanese EFL learners differ in using dictionaries". *Language Education and Technology* 40. 61-79.
- Koyama, T. and Takeuchi, O. (2004). "Comparing electronic and printed dictionaries: How the difference affected EFL learning". *JACET Bulletin* 38. 33-46.
- Koyama, T. and Takeuchi, O. (2007). "Does look-up frequency help reading comprehension of EFL learners? Two empirical studies of electronic dictionaries". *CALICO Journal* 25 (1). 110-125.
- Lew, R. (In press). "Online dictionaries of English". In Fuertes-Olivera, P. A. and Bergenholtz, H. (eds.). *E-Lexicography: The Internet, Digital Initiatives and Lexicography.* London and New York: Continuum.
- Longman Dictionary of Contemporary English. 5th ed. Harlow: Pearson Longman. 2009. [LDOCE5].
- Longman Dictionary of Contemporary English [online, 5th ed.]. http://www.ldoceonline.com/[Access date: 20. Feb. 2011].

- Nesi, H. (2000). "On screen or in print? Students' use of a learner's dictionary on CD-ROM and in book form". In Howarth, P. and Herington, R. (eds.). *EAP Learning Technologies*. Leeds: University Press. 106-114.
- Osaki, S. et al. (2003). "Electronic dictionary vs. printed dictionary: Accessing the appropriate meaning, reading comprehension and retention". In Murata, M., Yamada, S. and Tono, Y. (eds.). *Proceedings of ASIALEX '03 Tokyo*. Tokyo: Asialex. 205-212.
- Shizuka, T. (2003). "Efficiency of information retrieval from the electronic and the printed versions of a bilingual dictionary". *Language Education and Technology* 40. 15-33.
- Xu, X. (2010). "Study on the effect of dictionary use on second language incidental vocabulary acquisition: An empirical study of college English vocabulary learning strategy". *Journal of Language Teaching and Research* 1 (4). 519-523.

7. Appendix

e-LDOCE5



e-COBUILD6

