CrossMark

ReCALL **29**(3): 335–356. 2017 © European Association for Computer Assisted Language Learning
doi:10.1017/S0958344017000131
First published online 6 June 2017

Dictionary form in decoding, encoding and retention: Further insights

ANNA DZIEMIANKO

Adam Mickiewicz University in Poznań, Poland
(email: danna@wa.amu.edu.pl)

Abstract

The aim of the paper is to investigate the role of dictionary form (paper versus electronic) in language reception, production and retention. The body of existing research does not give a clear answer as to which dictionary medium benefits users more. Divergent findings from many studies into the topic might stem from differences in research methodology (including the various tasks, participants and dictionaries used by different authors). Even a series of studies conducted by one researcher (Dziemianko, 2010, 2011, 2012b) leads to contradictory conclusions, possibly because of the use of paper and electronic versions of existing dictionaries, and the resulting problem with isolating dictionary form as a factor. To be able to argue with confidence that the results obtained follow from different dictionary formats, rather than presentation issues, research methodology should be improved. To successfully generalize about the significance of the medium for decoding, encoding and learning, the current study replicates previous research, but the presentation of lexicographic data on paper and on screen is now balanced, and the paper/electronic opposition is operationalized more appropriately. A real online dictionary and its paper-based counterpart composed of printouts of screen displays were used in the experiment in which the meaning of English nouns and phrases was explained, and collocations were completed with missing prepositions. A delayed post-test checked the retention of the meanings and collocations. The results indicate that dictionary medium does not play a statistically significant role in reception and production, but it considerably affects retention.

Keywords: electronic dictionary, paper dictionary, decoding, encoding, retention

1 Introduction

Dictionaries in electronic form, in particular those available online, are becoming more and more popular. Some scholars go so far as to predict that "the days of most authoritative, monolingual print dictionaries may be numbered" (Nesi, 2015: 598). Even though electronic dictionaries are constantly gaining in popularity, they have not superseded paper ones in all parts of the world. The migration of lexicographic content to e-dictionaries of various kinds, however pronounced and irreversible it is, encourages reflection on the benefits of this shift to dictionary users, and language learners in particular. The conclusion is by no means obvious, as medium-specific features of electronic dictionaries are not really appreciated by users. Quite surprisingly, a comprehensive survey into users' demands from

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

online dictionaries conducted by Müller-Spitzer, Koplenig and Töpel (2011: 204–206; 2012: 446) shows that electronic dictionaries are expected to be, above all, reliable, clear and up-to-date. Thus, their most appreciated properties overlap with those of paper dictionaries. Features unique to dictionaries in electronic form, such as multimedia or adaptability, are of little consequence (Müller-Spitzer *et al*., 2011: 207).

If what is most appreciated in paper dictionaries should also characterize electronic ones, the fundamental question which suggests itself is whether dictionary form (medium) ultimately matters. The relative merits of both media (paper and electronic) have been widely discussed, and serious academic research has been devoted to the role of dictionary form. In a comprehensive review, Dziemianko (2012a) analyses over 20 experimental studies by different authors who, among other things, investigated the significance of dictionary form for language comprehension, production and learning. Unfortunately, this large body of research does not lead to any obvious conclusions; it is difficult to generalize from it about the effect that dictionary form has on language decoding, encoding and retention. There are many possible reasons for such inconclusiveness of previous research. The studies differ in aims, tasks, experimental design, scoring methods, target items and dictionaries. In particular, the dictionaries employed in the experiments had different content, interface, layout and typography. Reception, production and retention were operationalized in many different ways. Subjects represented various linguistic and cultural backgrounds or proficiency, and samples comprised from a few to several hundred participants.

That the question about the significance of dictionary medium for language comprehension, production and learning is by no means easy to answer can be clearly illustrated by a sequence of approximate replications conducted by Dziemianko (2010, 2011, 2012b).[1] In the studies in question, the same tests were taken in the same setting (Adam Mickiewicz University in Poznań, Poland) by participants representing the same proficiency in English (B2 to C1 in the Common Framework of Reference for Languages (CEFR)), linguistic and cultural background (Polish). However, in every experiment, the usefulness of paper and electronic versions of a different dictionary was tested: COBUILD6, LDOCE5 and OALDCE7, respectively. The dictionaries belong to the Big Five, the most reliable and renowned monolingual English learners' dictionaries (MLDs).

In each study, a pre-test, a main test and a retention test were employed, all of which were paper-based. In the main test, the participants worked with either the paper or electronic version of a given dictionary. Each participant performed a decoding task and an encoding one. In the former, the meaning of nine English nouns and phrases had to be explained (*backgammon, booby prize, clampdown, collateral damage, down under, dream ticket,*

[1] Approximate replications (or replications with changes) consist in repeating the original study with modifications introduced to one or more variables (e.g. tasks, population, setting) while securing comparability (Abbuhl, 2012: 298). Their aim is to determine the generalizability of the findings from the original investigation to the conditions created by the modification (Lindsay & Ehrenberg 1993: 217). Statistical significance is usually considered to provide sufficient grounds for concluding from research, because it suggests that the obtained results are unlikely to be a product of chance (and probably apply to the whole population), but it does not give an insight into their generality. To see how broadly the findings can be generalized beyond the original context, replication is necessary (Gast, 2009).

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

flapjack, onus, outcrop). In the latter, missing prepositions had to be supplied in English corpus-based sentences, which featured nine collocations (*on the blink, in cahoots with, up the creek, at gunpoint, wreak havoc on, in the offing, in the pipeline, under sedation, on the trot*).

The main test was immediately preceded by the pre-test, which consisted of the same receptive and productive tasks, except that the participants were requested to perform them without the help of any dictionary. On this basis, it was possible to identify (and exclude from further analyses) the cases where the participants knew the target phrases. The pre-test was complemented by a short questionnaire. The participants were requested to say whether they were familiar with advanced English learners' dictionaries and, if so, whether they used them on paper or in electronic form. This made it possible to determine whether their familiarity with this type of dictionary and experience in using either medium could be possible confounding factors.

The delayed retention test was conducted two weeks after the main test and was not announced in advance. It was the same as the pre-test except that the test items were reshuffled. Its aim was to check how many target meanings and collocations the participants remembered. In other words, it checked passive (receptive) recall and active (productive) recall, respectively.[2] Like in the pre-test, no dictionaries were allowed during the tasks.

Notwithstanding the same battery of tests, setting, participants' proficiency level as well as linguistic and cultural background, the three aforementioned approximate replications lead to different conclusions, briefly presented below.

In Study 1 (Dziemianko, 2010), 64 students of English used COBUILD6 either in book form (30 participants) or the online version available by a pin code to purchasers of the paper dictionary (34 participants). The results indicate that COBUILD6 online was significantly more useful in reception (98.7%), production (98.4%) and learning (passive recall 76.8%, active recall 63.9%) than COBUILD6 in book form (reception 92%, production 92.2%, passive recall 62.4%, active recall 45.6%; $p < 0.01$).[3]

In Study 2 (Dziemianko, 2011), 42 participants consulted LDOCE5 on paper and 45 participants were assigned to the free online version of the same dictionary. However, unlike in Study 1, dictionary form proved to have no significant effect on reception, production or the retention of meaning and collocations (reception: paper 95.8%, electronic 94.7%; production: paper 96.0%, electronic 95.4%; passive recall: paper 41.9%, electronic 43.5%; active recall: paper 33.3%, electronic 31.4%; $p > 0.05$).

In Study 3 (Dziemianko, 2012b), 42 participants worked with OALDCE7 in book form, 44 with the CD-ROM version of the same dictionary installed on PCs. Broadly speaking, the results confirm those obtained in the second study; dictionary medium proved to be inconsequential for language reception (paper 93.2%, electronic 96.1%), production (paper 95.2%, electronic 93.1%) and learning (passive recall: paper 34%, electronic 28.7%; active recall: paper 36.2%, electronic 23.8%; $p > 0.05$).

Overall then, in the case of COBUILD6, the electronic medium proved to be significantly more conducive to language reception, production and learning. By contrast, LDOCE5 online and OALDCE7 on CD-ROM were as useful in encoding, decoding and retention as the paper versions of these dictionaries.

[2] According to Nation (2001: 358–360) and Laufer and Goldstein (2004: 406), the retrieval of meaning is passive recall, while the retrieval of form is active recall.

[3] Unless clearly stated otherwise, percentages cited in the text refer to task success rates.

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

elasticity /ˌiːlæˈstɪsəti/ noun [U] **1** PHYSICS the ability of a substance to stretch easily and then return to its original shape quickly: *the strength and elasticity of silk* **2** the ability to change when the situation changes: *the elasticity of the labour market* **3** ECONOMICS the degree to which supply, DEMAND etc changes according to other economic conditions that change, for example the price of goods

Fig. 1. MEDAL2 entry for *elasticity*

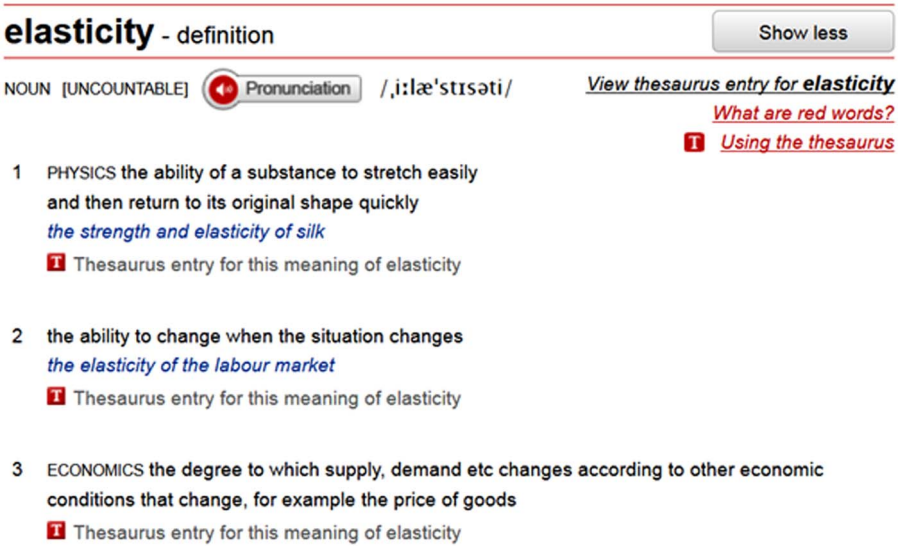


Fig. 2. MEDO entry for *elasticity*

Dziemianko (2012b) suggests many possible reasons for the divergent results, such as the layout of screen displays or noise levels – in particular intrusive, unsolicited adverts present in free online dictionaries (LDOCE5), but not in premium online versions (COBUILD6). Besides such inter-dictionary differences, reliance on existing dictionaries, which makes research more naturalistic, is bound to involve intra-dictionary factors which remain beyond the researcher's control. In fact, paper and electronic versions of the same dictionary are not directly comparable. A possible source of differences between the two formats of a dictionary might reside in structural indicators, both typographical (graphical aids such as font, colour, bold or italics) and non-typographical (asterisks, parentheses, icons and punctuation marks; see Gouws, 2003). Other possible confounding factors might include medium-specific information categories (e.g. multimedia, links to corpora or other dictionaries), layout (e.g. spacing between lines and textual condensation devices, see Corréard, 2002; Wiegand, 1996), hypertext and access facilities. To illustrate, Figures 1 and 2 provide an insight into differences between the online MEDO and print MEDAL2.

As can be seen, there are no codes in the e-dictionary, senses are given in new, numbered paragraphs which are clearly laid out, each example begins on a new line and is highlighted in blue, line spacing is wide, there are easily accessible audio files and hyperlinks to thesaurus entries. In MEDAL2, by contrast, the text is more condensed, senses and examples

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

do not start in new lines, line spacing is very narrow, there are no colours apart from black, and only capitals, italics and fonts distinguish information categories.

In electronic dictionaries, the outer access (i.e. the path leading the user to the desired headword) is largely simplified; arriving at the entry for the word boils down to typing it in the search window. In dictionaries which offer incremental search (see Lew, 2012a: 351) even this effort is superfluous, as it is enough to enter a few initial letters in the search widow and the dictionary suggests possible headwords. In print dictionaries, by contrast, there is no alternative to laborious page-turning (possibly helped by the thumb index and running heads) and scanning densely printed pages in order to finally land at the right headword. Further, in electronic dictionaries, advanced matching mechanisms help users extract multi-word expressions without the need to scroll through the whole entry. Instead, advanced matching makes a dictionary display possible matches of the target expression spanning a few words, not just the word typed into the box. In this way the user can select the expression needed from the list and immediately jump to it just by clicking, which automatically scrolls the dictionary page to the right section. Needless to say, there are no such shortcuts in paper dictionaries, where any multi-word expression which is not given headword status needs to be found within the entry for a constituent of that expression. As research shows, deciding which constituent is likely to include in its microstructure the desired multi-word expression can be a grave problem to dictionary users (e.g. Béjoint, 1981; Tono, 1987; Bogaards, 1990, 1991, 1992; Lew, 2012b).

Such differences between paper and electronic dictionary formats seriously hinder isolating the role of dictionary medium in research. It is clear that research methodology should be improved to enable researchers to successfully isolate dictionary form as a variable and convincingly argue that dictionary formats, and not presentation issues, account for the observed regularities. Aware of the limitations of her studies, Dziemianko (2012b: 211) recommends using printouts of the electronic dictionary screen display instead of existing paper dictionaries in further experiments. It appears that printout-based dictionary booklets can go a long way towards reducing typographical, non-typographical, layout and access disparities between paper and electronic dictionary formats.[4]

The aim of the present investigation is to further examine the effect of dictionary form on language reception, production and learning using an improved design in which the effect of the medium is more carefully isolated. This time, another MLD which belongs to the Big Five, MEDO, is used as test material. Interestingly, the editors of its paper counterpart, MEDAL2, have recently decided not to publish the dictionary in book form anymore.[5] This makes MEDO, a free online dictionary, perfectly suitable for research with improved methodology. By selecting yet another exemplar of the Big Five, the study also attempts to give an insight into the relative usefulness of MEDO, and thus extend the knowledge of the usefulness of MLDs gained from the previous replications. At this point, only the *Cambridge Advanced Learner Dictionary* is missing from the series of replications, with all other dictionaries included. A decision was made not to include this dictionary from the Big

[4] The electronic/paper opposition was operationalized in a similar way by Chen (2012: 226), who used a desktop dictionary and its printouts "so that factors related to dictionary information, lexicographical presentation, layout, font, and line spacing etc. are balanced off… to isolate the on-screen versus paper presentation."

[5] See Rundell (2013) for a concise justification.

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

Five in the current study to make the design of the experiment as close as possible to the previous ones, except for the improvements in methodology aimed at isolating dictionary form more successfully. For the same reason, the format that appears to be gaining in importance in Europe (i.e. mobile dictionaries, typically available via a smartphone app (or mobile web)) was not considered in the current replication, either. Even though in the foreseeable future the actual choices available to dictionary users may be different from print versus digital, the issue of how real or not the digital advantage is, and where it actually applies, remains highly topical.

The paper attempts to answer the following research questions:

1. Which dictionary form (paper or electronic) is more conducive to decoding, encoding and learning meanings and collocations?
2. What is the relative usefulness of MEDO for language reception, production and learning compared with COBUILD6, LDOCE5 and OALDCE7?

2 Methods

2.1 Materials

The tests designed by Dziemianko (2010) were used in the experiment. Thus, the main test consisted of decoding and encoding tasks. In the former, nine English words and phrases had to be explained in the participants' native language (Polish) or in English. In the latter, nine gapped corpus-based sentences were to be completed with the prepositions missing from collocations (see Section 1 for details). Participants did the tasks with the help of either MEDO or a purpose-built dictionary booklet, which consisted of colour printouts of MEDO screen displays. The existing paper counterpart, MEDAL2, was not used so as to reduce the influence of many confounding factors which could distort the effect of dictionary form (see Figures 1 and 2).

At the stage of compiling the booklet of printouts from MEDO screen displays, it was necessary to decide whether pages should be arranged alphabetically, like in a paper dictionary, or in the order which would coincide with the expected lookup sequence in MEDO. It was reasonable to assume that e-dictionary users would not bother about the alphabetical arrangement of the macrostructure, but would rather look up test items in the order of their appearance in the test sheet, which was not alphabetical. Thus the pages in the MEDO-based booklet were arranged in the same way as test items. This organization of the booklet macrostructure was to make the outer access as similar as possible to that of MEDO and save the participants finding target headwords in the alphabetical sequence, as e-dictionary users did not do this, either.

Another decision had to be made on the selection and arrangement of printouts in the situation where the desired collocation could be accessed in MEDO only by cross-references. First, *in the pipeline* and *up the creek* could be accessed from the entries for *pipeline* and *creek* through hyperlinks in the section titled PHRASES at the bottom of the entries or in the menu on the right. Second, *on the blink* and *on the trot* could not be found in the entries which MEDO opened by default after entering each base (*blink, trot*), because entries for verbs (*blink V, trot V*), not nouns, appeared first. It was thus necessary to follow the hyperlinks to noun entries or target collocations in the menu to the right of the verb entry. Finally, *wreak*

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

havoc on could be retrieved from the entry for *wreak*, but not *havoc*, which only gives a link to *wreak*.[6] In all these situations, the booklet imitated the MEDO access paths as closely as possible. A printout with the entry that the dictionary opened immediately after entering the base was given first (*pipeline, creek, blink V, trot V, havoc*). Then, a printout of the next step in the access path was provided (the collocation entries *in the pipeline, up the creek*, entries for *blink N, trot N, wreak*). In the case of *blink N* and *trot N*, the entries included hyperlinks to the target collocations, and the printouts of the collocation entries were supplied after those for the respective nouns.[7] In this way the access path to the desired collocation was as similar as possible for the users of the online dictionary and the booklet. Importantly, comments elicited from the participants after the test make it clear that the non-alphabetical arrangement of the booklet did not cause any confusion.

Apart from the main test, the original pre-test and post-test were used (Dziemianko, 2010). In the former, the participants' background knowledge of the target phrases and collocations was checked. The latter examined the retention of meaning and collocations. In both tests, the participants did the same tasks as in the main test without access to dictionaries. To reduce the learning effect, the arrangement of the test items was changed in each phase of the research.

2.2 Participants and procedures

Ninety-five Polish learners of English (B2-C1 in CEFR) took part in the experiment. First, they were asked to do the pre-test, which made it possible to identify the cases where a dictionary was not needed to perform the tasks. The pre-test was accompanied by a short questionnaire (Dziemianko, 2010) to investigate the participants' familiarity with dictionary types and formats. Immediately afterwards, the main test was administered. Each participant used either MEDO or the booklet created from MEDO printouts. Fifty-one participants were randomly assigned to work with the former, 44 with the latter. Access to the digital dictionary was provided in a computer lab, where the participants used desktop computers with uniform screens (resolution: 1920 × 1080, viewable size: 23 inches). As noted in Section 2.1, the booklet contained only the entries needed in the test, printed in colour on separate sheets of paper and arranged in the same order as the target items in the main test. Both paper dictionary and e-dictionary users were asked to do all the tasks one by one, in the

[6] Hausmann (2004, 2007) argues that for production purposes, collocations should be gathered in a dictionary under the semantically autonomous base (e.g. *cash* in the collocation *hard cash*), rather than the semantically dependent collocate (*hard*), whose meaning is contingent on the base (cf. *hard cash* vs *hard work*). Likewise, Lew (2012b) points out that when it is difficult to decide which word in a multi-word expression including a noun is less frequent, the expression should be treated in the noun entry. This matches users' expectations since, as research shows, they tend to prioritize nouns in their lookup choices when dealing with multi-word expressions (Béjoint, 1981; Bogaards, 1990; Lew, 2012b). However, in the entry for the noun *havoc*, MEDO gives only the following collocations: *cause/create havoc for* and *play havoc with*, and a cross reference to *wreak* at the bottom of the entry.

[7] As already noted, the verb entries for *blink* and *trot* offered hyperlinks to the target collocations in the menu on the right as well. Yet, it would be overoptimistic to assume that all users would instantly recognize the desired collocations among many other expressions in the menu. Besides, MEDO gives the hyperlinks to the collocations after those to the nouns *blink* and *trot*, which further reduces the chances of the collocations being selected first. The issue resurfaces in Section 4, where it is also illustrated in Figure 15.

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

order in which they appeared in the test sheets. Two weeks after the main test, the unexpected retention test was given. At each stage of the experiment, one participant did both the receptive and productive tasks, each of which consisted of nine target nouns and phrases (see Section 1). The *task* was thus a within-subject (repeated measures) factor. Dictionary *form* was a between-groups factor, because each participant had access to dictionary content either on paper or in electronic form.

The grading of the participants' answers in the production task was quite straightforward: one point was given for each correctly supplied preposition. Evaluating the receptive task was more complicated, as the participants could use their native language or English. To check the appropriacy of Polish equivalents, leading English-Polish dictionaries were consulted (*The New Kościuszko Foundation Dictionary* [NKFD], *Wielki Słownik Angielsko-Polski PWN-Oxford* [PWN Oxford]). Additionally, the four most popular bilingual dictionaries (Lew and Szarowska, 2017) available on the internet for free were referred to (Diki.pl, bab.la, Ling.pl, PONS). If the equivalent supplied by the participants in the reception test was present in any of them, a point was given. Descriptive explanations of meaning which the participants supplied in English were evaluated independently by the researcher and a native speaker of English, with very high inter-rater reliability (Pearson's correlation coefficient $r = 0.99$). Likewise, meaning explanations in Polish were assessed by the researcher and a native speaker of Polish. Both raters evaluated the responses identically ($r = 1.00$).

3 Results

3.1 The current replication

3.1.1 Dictionary-based lexical tasks. The questionnaire reveals that all the participants used MLDs. Moreover, their familiarity with dictionary formats was comparable in the MEDO group and in the booklet group (Table 1). This could not therefore be considered a factor accounting for differences in scores observed between these experimental conditions.

The results of the 2×2 repeated measures ANOVA (with *task* as a within-group factor and *form* as a between-groups factor) for the main test are given in Table 2.[8] Data on the participants' performance in the main test are presented in Figures 3–5.

Clearly, in the main test, the role of dictionary form was highly significant ($p = 0.01$). The test scores obtained with the assistance of MEDO were nearly 10% (97.7*100/89.2 = 109.5) better than those achieved after the consultation of the booklet (Figure 3). The effect size was big: over 34% of the between-subjects variance can be accounted for by dictionary form (Table 2).

Interestingly enough, neither task proved easier ($p = 0.65$); in production and reception, the participants' answers were correct in about 93% and 94% of all cases, respectively (Figure 4).

The interaction between dictionary form and task was not statistically significant, either ($p = 0.46$). In reception, the participants using the electronic dictionary were about 7% (97.2*100/91.2 = 106.6) more successful, while in production, the difference in favour of the electronic dictionary group approximated 13% (98.2*100/87.1 = 112.8, Figure 5).

[8] The analysis was conducted with the help of *Statistica* (GLM ANOVA procedure with default settings).

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

Table 1 *Participants' familiarity with dictionary formats*

	Groups		
	MEDO	booklet	*p* values (Z test for independent samples)
E-dictionaries	80%	73%	0.42
Paper dictionaries	4%	7%	0.52
Both dictionary formats	16%	20%	0.61

Table 2 *ANOVA results for the main test*

	Main test		
Factor	*F*	*p*	partial η^2
Form	8.29	0.01*	0.341
Task	0.21	0.65	0.013
Task × Form	0.57	0.46	0.034

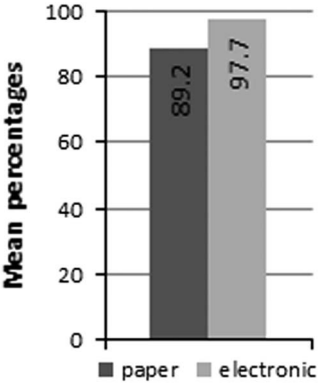


Fig. 3. Main test results by *form*

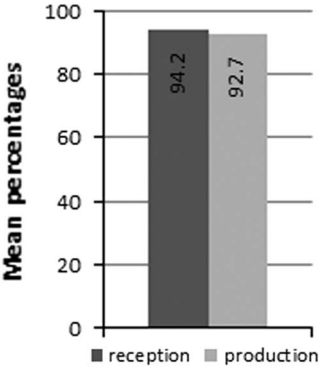


Fig. 4. Main test results by *task*

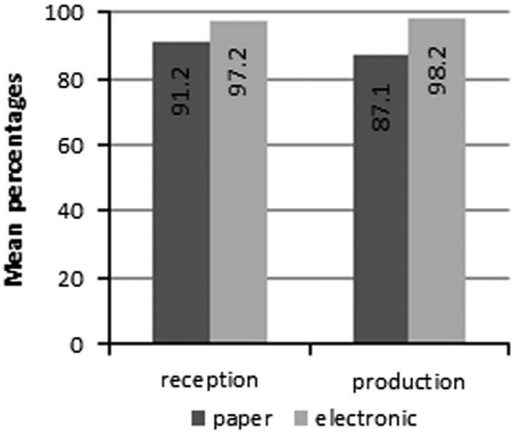


Fig. 5. Interaction *task x form* for the main test

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

Table 3 *ANOVA results for the retention test*

	Retention test		
Factor	*F*	*p*	partial η^2
Form	16.07	0.00*	0.501
Task	1.75	0.20	0.098
Task × Form	0.04	0.85	0.002

Neither of the differences was statistically significant on the Tukey HSD test (reception: $p = 0.50$; production: $p = 0.08$). The effect sizes of the interaction (3%) and *task* (1%, Table 2) were relatively small.

3.1.2 Retention. The 2x2 repeated measures ANOVA results for the retention test are collated in Table 3. The data are illustrated in Figures 6–8.

Significantly better retention results were obtained when the electronic dictionary had been used in the main test ($p = 0.00$, Table 3). About twice as much information (54.7/28.0 = 1.95) was remembered after reference to MEDO as after the consultation of the booklet (Figure 6). Over half of the between-subjects variance in retention can be explained by dictionary form (partial $\eta^2 = 0.501$).

The task did not prove to be as important for retention as dictionary form. Even though the retention of meaning was over one quarter (46.2*100/36.5 = 126.6) better than the retention of collocations (Figure 7), the difference was not statistically significant ($p = 0.20$).

The interaction effect was not significant, either ($p = 0.85$); the electronic medium had a positive and similar effect on passive and active recall. It is worth noting that over 87% (60.2*100/32.1 = 187.4) more words and phrases were correctly explained in the post-test after the consultation of MEDO than after reference to the booklet. Also, the electronic dictionary helped participants to learn over twice as many collocations (49.2*100/23.9 = 205.9) as its purpose-built paper counterpart (Figure 8). The Tukey HSD test showed both these differences to be statistically significant (passive recall: $p = 0.04$; active recall: $p = 0.03$). Overall, then, the retention of meaning and collocations was comparably (and significantly) enhanced by the electronic medium.

3.2 Four dictionaries compared

The use of the same battery of tests in all the replications (Dziemianko, 2010, 2011, 2012b, the current study) and the involvement of comparably proficient participants (B2-C1 in each study) with the same linguistic and cultural background make it possible to include one more factor, *dictionary*, and compare the results obtained in the main and retention tests in all the four experiments. *Dictionary* is a between-groups factor, because the participants in a study consulted only one dictionary. As there have been four replications so far, the factor has four levels, which correspond to the specific dictionaries used in the studies: COBUILD6, LDOCE5, OALDCE7, MEDO. The other factors taken into consideration are dictionary *form* (a between-groups factor; each participant consulted either the paper or

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

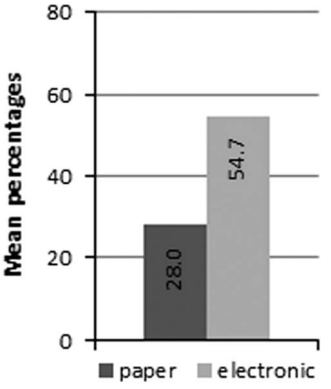


Fig. 6. Retention results by *form*

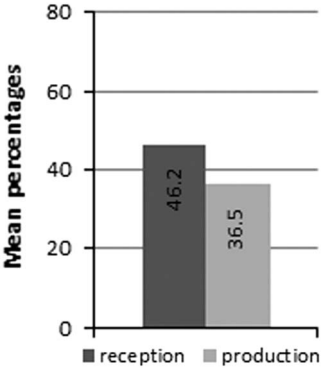


Fig. 7. Retention results by *task*

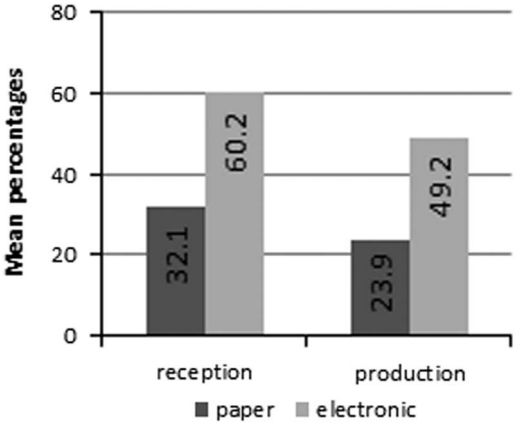


Fig. 8. Interaction *task* × *form* for retention

electronic version of a dictionary) and *task* (a within-subject factor; everyone did both tasks: receptive and productive). The results of the $4 \times 2 \times 2$ repeated measures ANOVA for the main test are presented in Table 4. Data on significant effects are illustrated graphically in Figures 9 and 10.

The ANOVA results indicate that the effect of dictionary form was statistically significant. Across all four studies, main test results based on electronic dictionary consultation were around 4% better than the results obtained with the help of paper dictionaries (96.5*100/92.8 = 103.9, $p = 0.00$). The significant two-factor interaction ($p = 0.01$) suggests, however, that the effect of dictionary form depended on the specific dictionary used in the research. Details on the post-hoc comparisons (Tukey HSD test) conducted for the significant interaction are given in Table 5.

The main test results supported by MEDO consultation (97.7%, Figure 10) were about one tenth better than those obtained with the help of the MEDO-based booklet (89.1%; 97.7*100/89.1 = 109.7), and the post-hoc comparisons indicate that the difference was statistically significant ($p = 0.01$). Also, the consultation of COBUILD6 online produced approximately one tenth better results than reference to the printouts from MEDO

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

(98.6*100/89.1 = 110.7), and the difference proved to be statistically significant as well ($p = 0.00$). No other differences had statistical significance on the Tukey HSD test. It is worth noting, then, that when all the dictionaries are taken into account, only the operationalization of dictionary *form* where the paper version was more parallel with the electronic format produced statistically significant results.

The 4x2x2 repeated measures ANOVA results for the retention test are shown in Table 6. Significant main and interaction effects are illustrated in Figures 11–14.

The ANOVA results indicate that the strongest effect on retention was exerted by the dictionaries consulted ($F = 14.57$, $p = 0.00$*, partial $\eta^2 = 0.406$). Over 40% of the between-subjects variance can be accounted for by the specific dictionaries used in the experiments. Most information was learnt with the help of COBUILD6; its users remembered over three fifths (62.2%) of the information from the main test two weeks later (Figure 11). The results of the post-hoc Tukey HSD test (Table 7) show that the retention results after the consultation of COBUILD6 were significantly better than those obtained when the other dictionaries had been used in the main test. Reliance on OALDCE7, LDOCE5 and MEDO produced comparable learning effects; from one third (OALDCE7: 30.7%) to about two fifths (LDOCE5: 37.5%; MEDO: 41.4%) of the information from the main test was retained.

Second, the ANOVA results show that, overall, electronic dictionaries were more conducive to learning than paper ones (Figure 12). The participants working with electronic

Table 4 *ANOVA results for the main test (four dictionaries)*

	Main test		
Factor	*F*	*p*	partial η^2
Dictionary	0.65	0.58	0.030
Form	9.81	0.00*	0.133
Dictionary × Form	3.80	0.01*	0.151
Task	0.08	0.77	0.001
Task × Dictionary	0.10	0.96	0.005
Task × Form	0.00	0.99	0.000
Task × Dictionary × Form	0.60	0.62	0.027

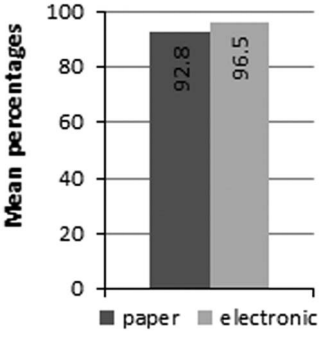


Fig. 9. Main test results by *form*

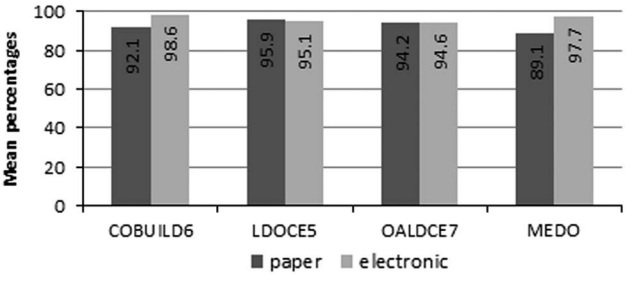


Fig. 10. Main test results *dictionary* × *form*

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

Table 5 *Tukey HSD test results for dictionary x form (main test, p values)*

		COBUILD6		LDOCE5		OALDCE7		MEDO	
Dictionary	Form	paper	electronic	paper	electronic	paper	electronic	paper	electronic
COBUILD6	paper		0.12	0.74	0.91	0.99	0.96	0.91	0.26
	electronic	0.12		0.95	0.81	0.57	0.70	0.00*	1.00
LDOCE5	paper	0.74	0.95		1.00	1.00	1.00	0.09	0.99
	electronic	0.91	0.81	1.00		1.00	1.00	0.20	0.95
OALDCE7	paper	0.99	0.57	1.00	1.00		1.00	0.40	0.79
	electronic	0.96	0.70	1.00	1.00	1.00		0.29	0.88
MEDO	paper	0.91	0.00*	0.09	0.20	0.40	0.29		0.01*
	electronic	0.26	1.00	0.99	0.95	0.79	0.88	0.01*	

dictionaries remembered nearly half of the information from the main test (47.2%), while among those who had had paper dictionaries at their disposal retention only approximated two fifths (38.7%). The difference in retention of 22% (47.2*100/38.7 = 122.0) in favour of electronic dictionary users was statistically significant ($p = 0.02$, Table 6). However, the statistically significant interaction between *dictionary* and *form* ($p = 0.00$, partial $\eta^2 = 0.192$, Table 6) suggests that the positive influence of the electronic medium on retention depended on the specific dictionary used. Table 8 presents the results of the Tukey post-hoc test for this interaction effect.

When both *dictionary* and *form* are taken into consideration, it turns out that there was a statistically significant effect of the medium on retention in the case of MEDO and its paper-based counterpart; about twice as much information was remembered when the e-dictionary had been used (54.7*100/28.0 = 195.3, $p = 0.01$, Figure 13).[9] Although the retention results obtained after the consultation of e-COBUILD6 exceeded by about one third those achieved on the basis of COBUILD6 on paper (70.3*100/54.0 = 130.2), the difference was not statistically significant ($p = 0.31$).[10] The participants who had consulted OALDCE7 and LDOCE5 in book form obtained, respectively, 33% (35.1*100/26.3 = 133.5) and 0.5% (37.6*100/37.4 = 100.5) better retention results than those who had used e-OALDCE7 and e-LDOCE5, but the effects were not statistically significant ($p > 0.05$). Clearly, the positive role of the electronic medium in stimulating learning, suggested by the statistically significant main effect of *form*, is indeed dictionary-specific. In reality, the beneficial effect of the electronic medium should be attributed to MEDO and, to a lesser extent, COBUILD6; but it was partly offset by OALDCE7 and LDOCE5.

Table 8 also shows that the consultation of the e-dictionaries influenced the effectiveness of learning to a different extent. COBUILD6 online (70.3%) contributed more to retention than the electronic (and paper) versions of OALDCE7 (26.3%) and LDOCE5 (37.4%) ($p = 0.00$ in all the cases). The learning based on COBUILD6 online (70.3%) and MEDO (54.7%) proved comparably successful; even though e-COBUILD6-assisted retention was over one

[9] Cf. Section 3.1.2.

[10] The prefixes *e-* and *p-* stand for *electronic* and *paper*, respectively, and are added to dictionary acronyms to differentiate between these two dictionary formats for present purposes, but are not part of dictionary titles.

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

Table 6 *ANOVA results for the retention test (four dictionaries)*

	Retention test		
Factor	*F*	*p*	partial η^2
Dictionary	14.57	0.00*	0.406
Form	5.72	0.02*	0.082
Dictionary × Form	5.07	0.00*	0.192
Task	9.80	0.00*	0.133
Task × Dictionary	0.94	0.43	0.042
Task × Form	0.16	0.69	0.003
Task × Dictionary × Form	0.16	0.92	0.007

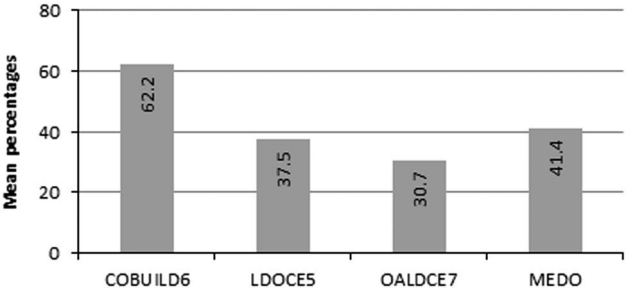


Fig. 11. Retention by dictionary

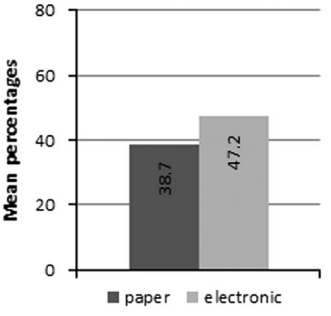


Fig. 12. Retention by *form*

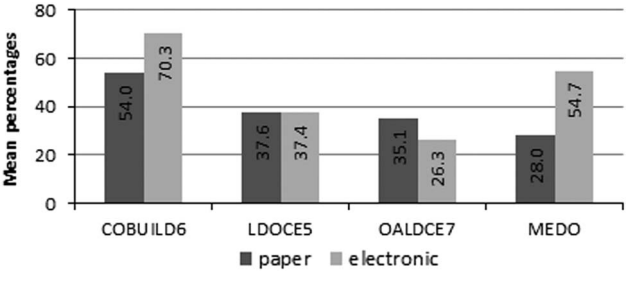


Fig. 13. Retention by *dictionary* and *form*

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

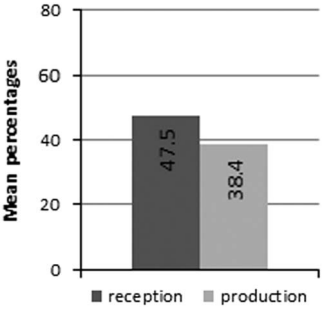


Fig. 14. Retention by *task*

fourth better (70.3*100/54.7 = 128.52), this effect was not statistically significant ($p = 0.37$). Retention based on MEDO (54.7%) was over twice as good as that based on e-OALDCE7 (26.3%), and the difference was statistically significant ($p = 0.00$). MEDO (54.7%) also stimulated nearly half again as much retention as LDOCE5 online (37.4%; 54.7/37.4 = 1.46), but the difference did not reach statistical significance ($p = 0.24$). Besides, learning after reference to MEDO was not statistically different from that which followed the consultation of any existing paper dictionary (COBUILD6, LDOCE5 or OALDCE7).

Finally, the ANOVA results in Table 6 show that the task largely affected retention ($F = 9.80$, $p = 0.00$* partial $\eta^2 = 0.133$). Passive recall was about one fourth more successful than active recall (47.50*100/38.40 = 123.70; Figure 14). Quite predictably, it was much easier for the participants to remember meaning than collocations.

4 Conclusions and discussion

In the light of these results, the answer to the first research question is only partly positive: the electronic version of MEDO was not significantly more beneficial to language reception and production than its paper counterpart. While MEDO users achieved better results in explaining meaning and supplying collocations than the booklet group, the differences were not statistically significant.[11] However, MEDO stimulated significantly better retention of meaning and collocations. Thus, for reception and production, the conclusions from the current study coincide with those from the two replications in which the paper and e-versions of LDOCE5 and OALDCE7 were comparably useful in these two tasks (Dziemianko, 2011, 2012b). For retention, in turn, the results resemble those obtained by Dziemianko (2010), where e-COBUILD6 proved to be much more useful in retention than COBUILD6 in book form.

Based on the outcomes, it appears that the digital format is more suitable for language learners. For one thing, the findings from the main test, though not statistically significant on either of the tasks separately, suggest that the digital format is better for immediate assistance in solving language problems; for another, the electronic format supports retention. These broad pedagogical recommendations should be further verified for

[11] As shown in Table 2, the effect of dictionary form in the main test was significant, but when the tasks were additionally taken into account, the differences in favour of MEDO in reception and production did not reach statistical significance.

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

Table 7 *Tukey HSD test results for dictionary (retention, p values)*

Dictionary	COBUILD6	LDOCE5	OALDCE7	MEDO
COBUILD6		0.00*	0.00*	0.00*
LDOCE5	0.00*		0.53	0.87
OALDCE7	0.00*	0.53		0.16
MEDO	0.00*	0.87	0.16	

Table 8 *Tukey HSD test results for dictionary × form (retention, p values)*

		COBUILD6		LDOCE5		OALDCE7		MEDO	
Dictionary	Form	paper	electronic	paper	electronic	paper	electronic	paper	electronic
COBUILD6	paper		0.31	0.31	0.29	0.15	0.01*	0.01*	1.00
	electronic	0.31		0.00*	0.00*	0.00*	0.00*	0.00*	0.37
LDOCE5	paper	0.31	0.00*		1.00	1.00	0.75	0.87	0.26
	electronic	0.29	0.00*	1.00		1.00	0.77	0.89	0.24
OALDCE7	paper	0.15	0.00*	1.00	1.00		0.92	0.97	0.12
	electronic	0.01*	0.00*	0.75	0.77	0.92		1.00	0.00*
MEDO	paper	0.01*	0.00*	0.87	0.89	0.97	1.00		0.01
	electronic	1.00	0.37	0.26	0.24	0.12	0.00*	0.01	

instructed and self-paced language learning, since the setting, context and comfort of dictionary use can be radically different in these two cases.

In the main test, the participants who used the MEDO-based booklet had the greatest problems with *on the blink, on the trot* and *wreak havoc on*. To some extent, the structure of this resource might be an explanation. As noted in Section 2.1, in the case of the first two collocations, the booklet gave entries for *blink* (V), *blink* (N), *on the blink,* and *trot* (V), *trot* (N), *on the trot*, to imitate the access structure in MEDO (Figure 15).

Although the hyperlinks were visible in printouts, fewer participants followed them. This resulted in scores approximating 80% (for both collocations) among the participants using the booklet, and 100% in the MEDO group. *Wreak havoc on* was problematic because the collocation could be extracted only from the example in the entry for *wreak*. The entry for *havoc*, likely to be looked up first (see the discussion in Section 2.1), gave a cross-reference to *wreak*. Apparently, the participants using the paper resource did not follow the cross-reference and stopped at *havoc*. Only about 47% of their answers were correct, as opposed to 90% in the other group.

Medium-specific features of online dictionaries are evidently difficult to transpose to paper; their paper-based images do not evoke similar reactions in users. In particular, hyperlinks fail to be perceived as cross-references when printed out.

The assessment of the relative usefulness of MEDO in comparison with COBUILD6, LDOCE5 and OALDCE7, conducted to answer research question two, indicates that MEDO was as useful for understanding meaning and supplying collocations as the paper and electronic versions of the other dictionaries. Quite similarly, the analysis of the post-test

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

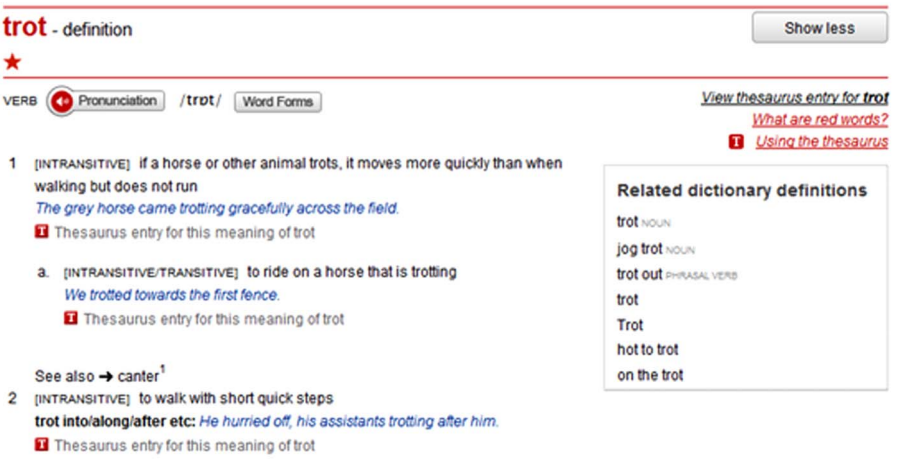


Fig. 15. Entry for *trot* (V) in MEDO

results reveals no statistically significant differences in retention after the consultation of MEDO and e-COBUILD6, p-COBUILD6, e-LDOCE5, p-LDOCE5 or p-OALDCE7. Only in the case of OALDCE7 on CD-ROM was there a statistically significant difference: reference to MEDO was much more conducive to retention than reliance on e-OALDCE7.

Interestingly, among e-dictionary users, the best retention results were obtained after reference to e-COBUILD6 (70.3%), followed by those achieved by the users of MEDO (54.7%), e-LDOCE5 (37.4%) and e-OALDCE7 (26.3%) (see Section 3.2). It might be worthwhile to reflect on possible reasons why the retention results varied so widely (though with the exception of the MEDO/e-OALDCE7 pair, not significantly) across the electronic dictionaries. To give an insight into dictionary content and interface, screenshots of the entry for *creek* from all the four dictionaries are presented in Figure 16. This headword was selected for comparison because differences in the retention results obtained for *up the creek* were the greatest (e-COBUILD6 63%, MEDO 26.7%, e-LDOCE5 4.8%, e-OALDCE7 0%).

In all these dictionaries, the target phrase occupies third place, whether numbered (e-LDOCE5, e-COBUILD6) or not (e-OALDCE7 and MEDO). The definitions in e-LDOCE5, e-OLDCE7 and MEDO are virtually the same: *(to be) in a (very) difficult (or bad) situation*. Only e-COBUILD6 offers a full-sentence definition. It is also the only dictionary that does not give any example of usage, probably because the contextual definition shows how to use the phrase and clearly indicates the appropriate context: *If someone is up the creek, they are in a bad or difficult situation, or are wrong in some way*.[12] The examples in the other dictionaries are very similar; all of them indicate the unfavourable situational context which justifies the use of the target phrase (*without my car* – e-OALDCE7; *If I don't get my passport by Friday* – e-LDOCE5; *If we don't get this contract* – MEDO). Thus, dictionary content does not seem to be a possible reason for the better retention results obtained for *up the creek* after the consultation of e-COBUILD6 and

[12] The exemplifying function of COBUILD definitions was discussed by Hanks (1987: 135), who noted that they are not only statements about word meaning, but also about word use.

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

MEDO than after reference to e-OALDCE7 or e-LDOCE5. Possibly, the presentation of lexicographic data might hold a clue.

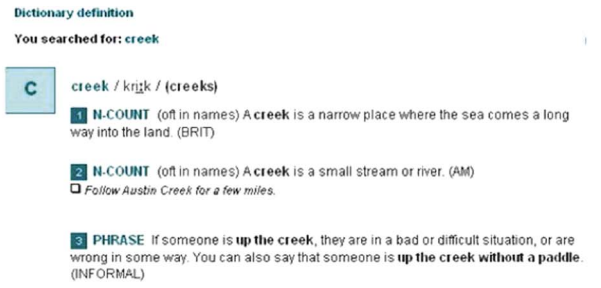
As can be seen in Figure 16, MEDO and e-COBUILD6 are much more sparing with colours than the other dictionaries. MEDO gives entry content in black (definitions, phrases/collocations, headword, part of speech, syntactic classification, transcription) and blue (examples) only. Additionally, thesaurus and pronunciation icons are in red, reference to thesaurus entries – in grey. E-COBUILD6 does not employ many colours, either, only black (definitions, examples, transcription, word forms, regional and register labels) and turquoise (grammatical labels, sense numbers, headword). By contrast, e-LDOCE5 and e-OALDCE7 ooze with colours. Dark blue (headword), cobalt (collocations), turquoise (word forms and loudspeaker icons), red (synonyms, idioms, cross-references), green (regional and register labels) and black (definitions, examples, transcription, part of speech) can be easily seen in the e-OALDCE7 entry. Red (sense numbers), crimson (related topics), purple (regional and style labels), green (syntactic classification), cobalt (headword, collocations) and black (examples) feature in e-LDOCE5.[13] Admittedly, colours have been found to significantly improve retention after dictionary consultation (Dziemianko, 2015). Yet, in that experiment, immediate (rather than delayed) retention was tested, and only labels were given in colour (part-of-speech labels – green; syntax labels – blue; style labels – dark red); the other entry components were in black. The much more colourful entries in e-LDOCE5 and e-OALDCE7 demonstrate considerable highlighting density, i.e. the extent to which highlighting is employed. In other words, an overall bigger proportion of entry content is highlighted. However, this highlighting density does not correlate with highlighting validity, or the probability that target information (relevant to the task at hand) is highlighted (see Tamborello & Byrne, 2007: 184). As pointed out above, more information categories were highlighted in different colours in those two dictionaries than in MEDO or e-COBUILD6, and many of them were not relevant to the task in the experiment. Yet, as colours do increase visual salience and affect search (Tamborello & Byrne, 2007: 182), the participants might have simply been too distracted by them. It is possible that the amount of colour increases extraneous cognitive load, which results from non-optimal data presentation. When extraneous cognitive load is raised, germane cognitive load (associated with processing the internal complexity of the information to be learnt) decreases, because more working memory resources are wasted on coping with the non-optimal data presentation, and fewer are engaged in processing the target content. Unfortunately, such unwelcome shift in the deployment of cognitive powers impedes learning (Sweller, Van Merrienboer & Pass, 1998: 263–264; Sweller, 2010: 124). Thus, the entries from e-LDOCE5 and e-OALDCE7 may imply that profuse highlighting in colour can absorb too much of users' working memory, and turn out to be more of a distraction than a help in learning. Yet, based on the analysis of one case (*up the creek*), this remark is only a hypothesis which should be verified in further research. It is also worth noting that COBUILD6 is the only dictionary where the bare *up the creek* is highlighted. This might have been another reason for the participants' greater success in dealing with the target phrase.

The current study suggests that it is necessary to develop still better ways of comparing the usefulness of paper and electronic dictionaries. While using printouts of online entries

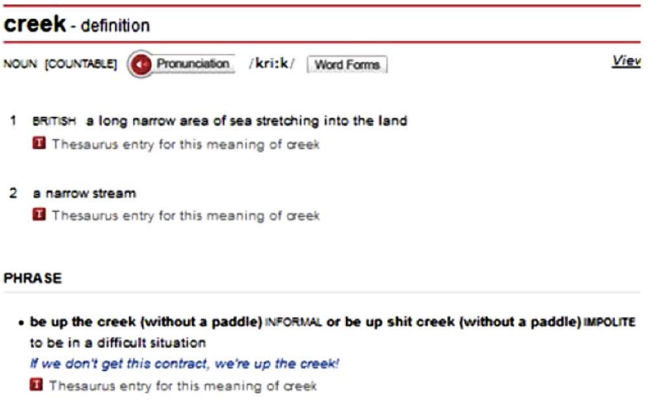
[13] This abundance of colours seems to be at odds with Almind's (2005: 39) recommendation for sparing and discreet use of colours in dictionaries to highlight only key elements of the entry without disturbing its legibility.

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

COBUILD6



MEDO



LDOCE5



OALDCE7

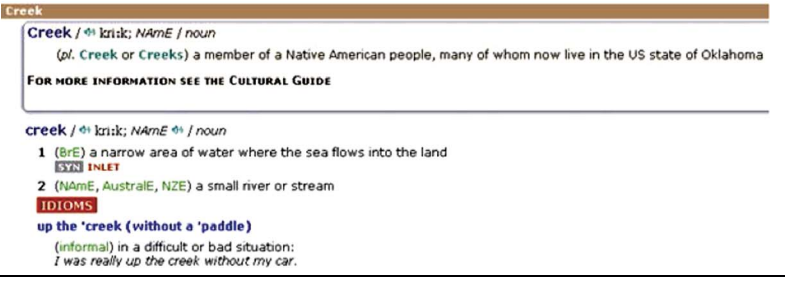


Fig. 16. COBUILD6, LDOCE5, OALDCE7 and MEDO entries for *creek*

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

instead of existing paper dictionaries is a step in the right direction, the paper and electronic media need to be made even more directly comparable. Test items should be selected so that the use of medium-specific features (such as hypertext or multimedia) would not be required. Also, the entries for the desired part of speech should be immediately available to save the participants from browsing different entries to find the right one. It might even be necessary to fabricate microstructures (both online and on paper) to make them exactly comparable, and in this way isolate the role of dictionary medium more successfully. Is seems, though, that this would entail a considerable loss of naturalness, which is also a limitation of the current study. Hopefully, further experiments will make it possible to come still closer to an answer to the question as to which dictionary format benefits users more.

References

Dictionaries

[COBUILD6] Sinclair, J. (ed.) (2008) *Collins COBUILD advanced English dictionary* (6th edn.). Boston: Heinle Cengage Learning; Glasgow: Harper Collins. http://www.mycobuild.com

[LDOCE5] Mayor, M. (ed.) (2009) *Longman dictionary of contemporary English* (5th edn.). Harlow: Longman. http://www.ldoceonline.com

[MEDAL2] Rundell, M. (ed.) (2007) *Macmillan English dictionary for advanced learners* (2nd edn.). Oxford: Macmillan Education.

[MEDO] *Macmillan English dictionary online*. http://www.macmillandictionary.com

[NKFD] Fisiak, J. (ed.) (2003) *The new Kościuszko foundation dictionary*. Kraków: Universitas.

[OALDCE7] Wehmeier, S. (ed.) (2005) *Oxford advanced learner's dictionary of current English* (7th edn.). Oxford: Oxford University Press.

[PWN Oxford] Linde-Usiekniewicz, J., Lewandowska-Tomaszczyk, B., Fisiak, J., Piotrowski, T., Beręsewicz, P. and Anioł, B. (eds.) (2005) *Wielki słownik angielsko-polski PWN-Oxford*. Warszawa: Wydawnictwo Naukowe PWN.

bab.la http://pl.bab.la/

Diki.pl https://www.diki.pl/

Ling.pl https://ling.pl/

PONS http://pl.pons.com/

Other references

Abbuhl, R. (2012) Why, when, and how to replicate research. In: Mackey, A. and Gass, S. M. (eds.), *Research methods in second language acquisition: A practical guide*. Oxford: Blackwell, 296–312.

Almind, R. (2005) Designing internet dictionaries. *Hermes*, **34**: 37–54.

Béjoint, H. (1981) The foreign student's use of monolingual English dictionaries: A study of language needs and reference skills. *Applied Linguistics*, **2**(3): 207–222.

Bogaards, P. (1990) Où cherche-t-on dans le dictionnaire? *International Journal of Lexicography*, **3**(2): 79–102.

Bogaards, P. (1991) Word frequency in the search strategies of French dictionary users. *Lexicographica*, **7**: 202–212.

Bogaards, P. (1992) French dictionary users and word frequency. In: Tommola, K. H., Varantola, K., Salami-Tononen, T. and Schopp J. (eds.), *EURALEX '92 proceedings. Papers submitted to the 5th EURALEX International Congress*. Tampere: Department of Translation Studies, University of Tampere, 51–59.

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

Chen, Y. (2012) Dictionary use and vocabulary learning in the context of reading. *International Journal of Lexicography*, **25**(2): 216–247.

Corréard, M.-H. (2002) Are space-saving strategies relevant in electronic dictionaries? In: Braasch, A. and Povlsen, C. (eds.), *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 12–17, 2002*, Vol. 2. Copenhagen: Center for Sprogteknologi, Copenhagen University, 463–470.

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. (2011) Does dictionary form really matter? In: Akasu, K. and Uchida, S. (eds.), *ASIALEX2011 proceedings. Lexicography: theoretical and practical perspectives*. Kyoto: Asian Association for Lexicography, 92–101.

Dziemianko, A. (2012a) On the use(fulness) of paper and electronic dictionaries. In: Granger, S. and Paquot, M. (eds.), *Electronic lexicography*. Oxford: Oxford University Press, 319–341.

Dziemianko, A. (2012b) Why one and two do not make three: Dictionary form revisited. *Lexikos*, **22**: 195–216.

Dziemianko, A. (2015) Colours in online dictionaries: A case of functional labels. *International Journal of Lexicography*, **28**(1): 27–61.

Gast, D. L. (2009) *Single subject research methodology in behavioral sciences*. New York: Routledge.

Gouws, R. (2003) Types of articles, their structure and different types of lemmata. In: Van Sterkenburg, P. (ed.), *A practical guide to lexicography*. Amsterdam: John Benjamins, 34–43.

Hanks, P. (1987) Definitions and explanations. In: Sinclair, J. M. (ed.), *Looking up: An account of the COBUILD project in lexical computing*. London: Harper Collins, 116–136.

Hausmann, F. J. (2004) Was sind eigentlich Kollokationen? In: Steyer, K. (ed.), *Wortverbindungen: Mehr oder weniger fest*. Berlin: De Gruyter, 309–334.

Hausmann, F. J. (2007) Die Kollokation im Rahmen der Phraseologie: Systematische und historische Darstellung. *Zeitschrift für Anglistik und Amerikanistik*, **55**(3): 217–234.

Laufer, B. and Goldstein, Z. (2004) Testing vocabulary knowledge: Size, strength and computer adaptiveness. *Language learning*, **54**: 399–436.

Lew, R. (2012a) How can we make electronic dictionaries more effective? In: Granger, S. and Paquot, M. (eds.), *Electronic lexicography*. Oxford: Oxford University Press, 343–362.

Lew, R. (2012b) The role of syntactic class, frequency, and word order in looking up English multi-word expressions. *Lexikos*, **22**: 243–260.

Lew, R. and Szarowska, A. (2017) Evaluating online bilingual dictionaries: The case of popular free English-Polish dictionaries. *ReCALL*, **29**(2): 138–159.

Lindsay, R. M. and Ehrenberg, A. S. C. (1993) The design of replicated studies. *The American Statistician*, **47**(3): 217–228.

Müller-Spitzer, C., Koplenig, A. and Töpel, A. (2011) What makes a good online dictionary? Empirical insights from an interdisciplinary research project. In: Kosem, I. and Kosem, K. (eds.), *Electronic lexicography in the 21st century: New applications for new users. Proceedings of eLEX2011*. Trojina: Institute for Applied Slovene Studies, 203–208.

Müller-Spitzer, C., Koplenig, A. and Töpel, A. (2012) Online dictionary use: Key findings from an empirical research project. In: Granger, S. and Paquot, M. (eds.), *Electronic lexicography*. Oxford: Oxford University Press, 425–457.

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

Nesi, H. (2015) The demands of users and the publishing world: Printed or online, free or paid for? In: Durkin, P. (ed.), *The Oxford handbook of lexicography*. Oxford: Oxford University Press, 579–589.

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press

Rundell, M. (2013) *Macmillan dictionary: Our move from print to online – your questions answered*. http://www.youtube.com/watch?v=PGK4QTDsA_Q&list=PLbEWGLATRxw9oBTqANj69Pz2RIHFoo3bR&index=3.

Sweller, J. (2010) Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review*, **22**(2): 123–138.

Sweller, J., Van Merrienboer, J. and Pass, F. (1998) Cognitive architecture and instructional design. *Educational Psychology Review*, **10**(3): 251–296.

Tamborello, F. and Byrne, M. (2007) Adaptive but non-optimal visual search behavior with highlighted displays. *Cognitive Systems Research*, **8**: 182–191.

Tono, Y. (1987) *Which word do you look up first? A study of dictionary reference skills*. Tokyo Gakugei University, Tokyo, unpublished M.Ed. dissertation.

Wiegand, H. E. (1996) Textual condensation in printed dictionaries: A theoretical draft. *Lexikos*, **6**: 133–158.

https://doi.org/10.1017/S0958344017000131 Published online by Cambridge University Press