Warnings in Learners' Dictionaries: Do They Help to Correct Errors and Learn Usage?

Anna Dziemianko
danna@amu.edu.pl
Adam Mickiewicz University in Poznań, Poland

ABSTRACT

The paper aims to investigate the effect of warning messages in monolingual English learners' dictionaries on the accuracy of error correction as well as immediate and delayed retention of usage. An attempt is also made to see whether the position of information useful for task performance in entries affects error correction and learning. Two types of warnings are investigated: deduction-oriented warning boxes and induction-oriented standalone examples of errors. In an online experiment, 162 upper-intermediate learners of English corrected errors in 18 sentences with the help of purpose-built dictionary entries. Three types of entries were created: with boxed warnings, with standalone examples of errors, and without any warnings. In all of them, the distribution of information relevant to task performance was strictly controlled; it was placed in entry initial, medial and final positions. Their results show a significant and positive role of warnings in monolingual learners' dictionaries (MLDs). Both warning types considerably improve error correction accuracy as well as immediate and delayed retention of usage, with warning boxes outdoing standalone examples of errors in all three respects. Entries without any warnings are the least effective. The position of relevant information in entries proves to be inconsequential for error correction or learning usage. The paper argues for the inclusion of warnings in monolingual English learners' dictionaries and suggests that they be better adjusted to the needs of speakers from specific mother tongue backgrounds.

Keywords: warnings; examples of errors; learners' dictionaries; error correction; learning

INTRODUCTION

A key objective of monolingual English learners' dictionaries (MLDs) is to help language learners in their encoding activities, including writing in academic or professional environments. Research shows that dictionary-based language production hinges on examples; it is examples that users most often turn to in search for relevant encoding information (Dziemianko, 2006). Learners themselves openly appreciate examples in dictionaries and admit they would like to see more of them (Farina, 2019: pp. 469-470). However, some examples might be unsuitable for language production, as dictionaries often fail to distinguish between decoding examples, which facilitate comprehension, and encoding examples, which help in production (Frankenberg-Garcia, 2015, p. 493).

On the other hand, appropriately curated corpus examples were repeatedly proved quite useful for error correction and avoidance. Frankenberg-Garcia (2012) concluded that such examples help learners to rectify common learner errors in English sentences much more than

eISSN: 2550-2131
ISSN: 1675-8021

definitions, and multiple examples are even more beneficial. In a conceptual replication, where error-prone Portuguese sentences had to be translated into English, Frankenberg-Garcia (2014) found that only multiple corpus examples in entries help to prevent errors, but not single ones, which were only as good as definitions. However, her further research suggests that increasing the number of relevant examples does not guarantee success in language production. In an experiment where participants had to revise their own translations for possible errors, Frankeberg-Garcia (2015) observed that multiple corpus examples which exhibit the target syntax are no better than one such example in the entry.

Importantly, Frankenberg-Garcia (2012, 2014, 2015) did not employ examples from existing MLDs in her studies, but judiciously selected corpus ones. They were deliberately biased in favor of the errors in the tests and included precisely the collocation and colligation patterns that were required in the production tasks. Hand-picked multiple examples ensured reiterated exposure to target word usage to facilitate detecting conventional patters of use. Yet, as noted above, actual dictionary examples may fail to address specific language production difficulties and prove irrelevant to the task at hand. Besides, considering the limited presentation space, even online MLDs usually cannot afford to give a few examples showing exactly the same pattern of use in one entry. Admittedly, extra corpus examples are sometimes provided in expandable boxes clickable on demand (called *Extra Examples* in the *Oxford Advanced Learner's Dictionary* (OALD) and *More examples* in the *Cambridge Advanced Learner's Dictionary* (CALD)), special entry sections located at the end of the entry (labeled *Examples from the Corpus* in the *Longman Dictionary of Contemporary English* (LDOCE)), or on a separate page (titled *Sentences* in the *Collins Online Dictionary* (COD)). Yet, learners may be unwilling to access such additional examples in real-life situations, as they want to find the needed information quickly, and preferably in the first place they look (Chan, 2012, p. 87). Besides, extra examples are either arranged by sense or not sorted at all (Frankenberg-Garcia, 2014, p. 141). As lexico-grammatical patters are not a grouping criterion, retrieving relevant information about word valence might be daunting. The challenge might be compounded by the fact that such examples may be jumbled up with no distinction between the parts of speech of the headword (cf. examples of *challenge* as a noun and a verb muddled up in one place in COD).[1]

It is not surprising, then, that more research is called for to develop efficient ways of making encoding examples accessible to learners (Frankenberg-Garcia, 2015, p. 507). However, the pertinent questions seem to be not only *how* to include them, but also *what* to include. After all, dictionary examples "are not necessarily geared to language production errors and certainly do not provide repeated exposure to specific target structures that can be problematic to learners with a specific mother tongue background" (Frankenberg-Garcia, 2012, p. 287). Possibly, introducing warning messages alerting users to grammatical errors might offer a solution.

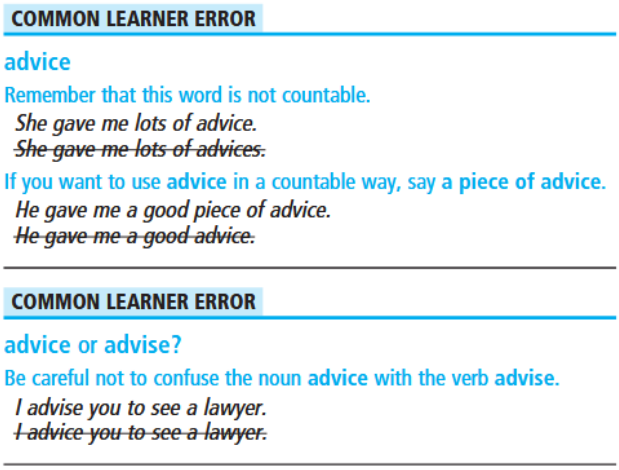
In MLDs, warnings usually take the form of boxes or notes, which often include examples of typical learner errors.[2] Dictionary warnings alert users to recurrent issues identified on the basis of learner error analysis. To competently identify typical errors, a reliable learner corpus is needed which sources texts produced by language learners at each main proficiency level and from a variety of L1 backgrounds. Such a database makes it possible to detect those learner errors which are frequent and widespread enough to be worth addressing in MLDs. A way to do it is to include

[1] Admittedly, Ptasznik (2024) found that advanced dictionary users are quite adept at extracting the necessary information from supplementary corpus examples. Yet, his participants did not have to scroll down the page, click a box or open a different page to access extra examples; they were spoon-fed with them in test materials.

[2] Osada, Sugimoto, Asada & Komuro (2015, p. 43) estimate that about one fifth of such notes in a learners' dictionary do not give examples of errors.

eISSN: 2550-2131
ISSN: 1675-8021

warning messages, like "Common Learner Error" notes in CALD, "Get It Right" in the *Macmillan Dictionary English* app (MDE), "Grammar" boxes in LDOCE, or untitled pink-shaded boxes in OALD (see Figures 1-4).[3]


COMMON LEARNER ERROR

advice

Remember that this word is not countable.

She gave me lots of advice.

~~*She gave me lots of advices.*~~

If you want to use **advice** in a countable way, say a **piece of advice.**

He gave me a good piece of advice.

~~*He gave me a good advice.*~~

COMMON LEARNER ERROR

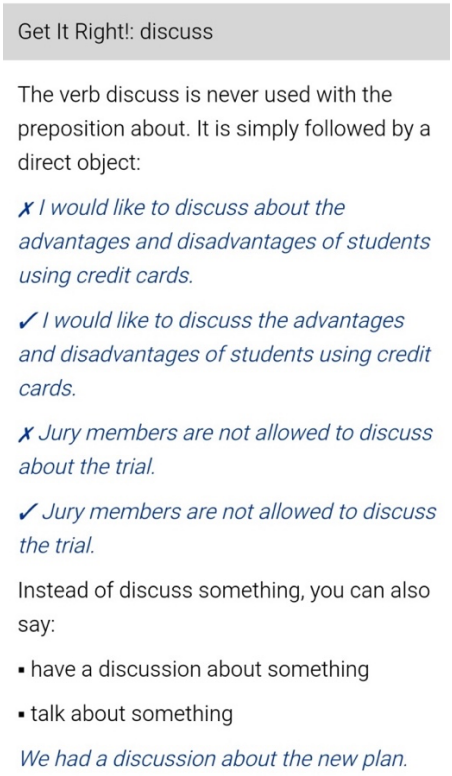
advice or advise?

Be careful not to confuse the noun **advice** with the verb **advise.**

I advise you to see a lawyer.

~~*I advice you to see a lawyer.*~~

FIGURE 1. A "Common Learner Error" box in CALD


Get It Right!: discuss

The verb discuss is never used with the preposition about. It is simply followed by a direct object:

✗ *I would like to discuss about the advantages and disadvantages of students using credit cards.*

✓ *I would like to discuss the advantages and disadvantages of students using credit cards.*

✗ *Jury members are not allowed to discuss about the trial.*

✓ *Jury members are not allowed to discuss the trial.*

Instead of discuss something, you can also say:

- have a discussion about something
- talk about something

We had a discussion about the new plan.

FIGURE 2. A "Get It Right" box in MDE (smartphone app)

[3] Unless clearly stated otherwise, free online versions of the MLDs are meant. The *Macmillan English Dictionary* website closed on 30th June, 2023 after 14 years of online excellence. Still, the content is available on the smartphone app *Macmillan Dictionary English* (MDE).

eISSN: 2550-2131
ISSN: 1675-8021

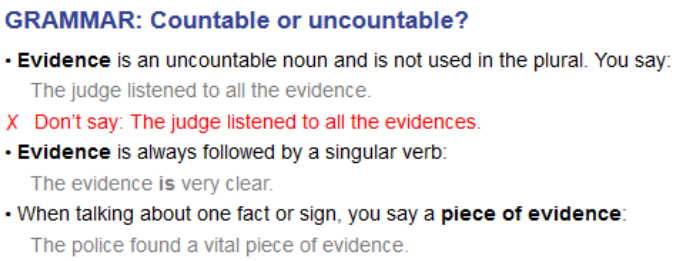


FIGURE 3. A "Grammar" box in LDOCE

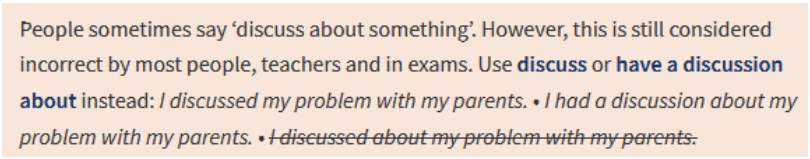


FIGURE 4. A warning message in OALD

The warning boxes in the main MLDs differ in formatting, but they share content components. A warning box opens with a description of a reason for the error and a clear explanation of how to use (or not to use) the target word. This makes it deduction-oriented.[4] The explanations are followed by examples of correct and incorrect usage. In all dictionaries but MDE, examples of correct usage precede examples of errors. Except for LDOCE, all examples are italicized. Examples of errors are always additionally highlighted – printed in ~~*striketrough font*~~ in CALD and OALD, marked by a cross (*X*) in MDE, and a red cross (*X*) and a *Don't say* note in LDOCE.

Interestingly, dictionary use is considered deduction-oriented, too (Tsai, 2019); dictionaries first give explicit rules and spell out lexico-grammatical patterns, and then illustrate them with examples. Thus, warning boxes reflect the typical arrangement of entry content (rules first, examples next). Yet, online MLDs cannot expound *all* the patterns and rules, only the most common ones. Others are just illustrated in corpus examples. To find out more about them, learners are encouraged to draw their own conclusions about language use from linguistic evidence. This do doubt involves induction, or the exploration of multiple language exemplars to generalize rules and patterns (Tsai, 2019, p. 808).

It is worth pointing out that corpus examples in entries reveal structures which are permissible, but not those which are unacceptable. Likewise, they do not alert users to the fact that some combinations are impossible with some words, even though they make perfect sense with others. That evidence-driven learners' conclusions concerning word usage are error-prone was obvious to the forefather of learners' dictionaries, A. S. Hornby, who long ago recognized the risk of production errors due to false analogies:

> "the learner … may suppose that because he has heard or seen "I intend (want, propose) to come," he may say or write "I suggest to come" [...] Because "He began talking about the weather" means about the same as "He began to talk about the weather", the learner may suppose, wrongly of course, that "He stopped talking about the weather" means the same as "He stopped to talk about the weather"
>
> (Hornby, 1956, p. v)

[4] In the deductive (rule-driven) pedagogical approach, learners are directly given prescribed rules and patterns of language (Tsai, 2019, p. 808).

eISSN: 2550-2131
ISSN: 1675-8021

False analogies may result not only from such second language generalization, but also from incongruence between a user's native language and English. To illustrate, words that in some languages may be in the plural have English equivalents which are singular only (e.g., French *nouvelles* and the English singular noun *news*). No wonder, then, that the best dictionary for learners is believed to be one that would supply guidance on syntax together with advice on pitfalls to avoid (Béjoint, 1994, p. 210).

It is dictionary warnings that caution users against linguistic pitfalls. They exemplify errors which are typical for language learners and alert users to frequent problems with language use. However, there are important issues related to their inclusion in MLDs. First, they may not live up to advanced learners' expectations. In fact, they surprisingly often do not address problems relevant to the advanced level (De Cock & Granger, 2004, p. 82). Second, although syntactic anisomorphism between languages can negatively affect linguistic performance, it is not accentuated in MLD warnings. MLDs represent a one-size-fits-all model; they do not cater for the needs of a particular language community, but address cross-linguistic problems typical for speakers of different L1s worldwide. Consequently, the errors they point out may not reflect the specific issues with English that speakers of a given L1 have. Third, it is not clear how to make warnings perceptually prominent. Visual enhancement techniques can take different form, e.g., CAPITALS, **bold print**, ~~strikethrough font~~, *italics*, underlining, font color, background color, border. The warnings in Figures 1-4 reveal a wide variety of highlighting methods adopted to increase noticeability, which is considered a prerequisite for successful dictionary lookup (Nural, Nesi & Çakar, 2022). Unfortunately, no justification is ever provided for the specific techniques used. Nural Nesi and Çakar (2022) observe that even in one dictionary (LDOCE) a range of typographical enhancements are employed to bring out warning notes, but their implementation seems entirely haphazard.

Fourth, the prominence of lexicographic information can be affected by its place in the entry. Empirical research into the findability of lexicographic information reveals entry tops to be the most salient entry parts. It is the beginning of the entry that users tend to focus on (and often fail to go beyond), even when what they find obviously does not fit the context at hand (e.g., Chen, 2017). However, there is also research which points to entry ends as the most salient (Dziemianko, 2014; Nesi & Tan, 2011). In an attempt to account for such divergent results, it is conjectured that advanced learners (and experienced dictionary users) might have got used to finding the most common information at the top of the entry and choose to look directly at the entry's bottom in search of what is more advanced and less obvious in the language. In this context, the study by Dziemianko (2014) requires attention, as it involves boxed dictionary information. It investigated the effect of the positioning and presentation of collocations on their use and retention. The advantage of entry-final position was observed only when collocations were highlighted in bold, but not when they were grouped in boxes. In the latter case, the success of extracting collocations was comparable when the boxes were given at the beginning of the entry and its end. This might be an interesting observation for the current study; the role of the positioning of warnings in entries might depend on whether they are in boxes or not.

This short overview of the rationale for warnings in MLDs, their form and appearance as well issues related to their inclusion does not reveal whether they are indeed useful for language learners. This is what the next section explores.

eISSN: 2550-2131
ISSN: 1675-8021

LITERATURE REVIEW

Not much is known about the relative usefulness of warnings in MLDs. In a paper-based study by Chan (2012), sentence grammaticality was judged with the help of entries from the *Cambridge Advanced Learner's Dictionary* (3rd edition, CALD3), some of which included warning boxes. The results show that participants neglected the boxes in decision making. Only in about 20% of their decisions did they use them. By contrast, examples and definitions informed as much as 80% and 50% of their judgements, respectively. However, warning boxes were not present in all the entries offered for consultation; only four out of ten CALD3 microstructures happened to include them.

More recently, Nural, Nesi and Çakar (2022) explored warning messages in LDOCE. A thorough metalexicographic analysis allowed them to identify five categories of warnings in the dictionary, based on formatting features. Yet, no systematic connection between the categories and either error types or enhancement techniques was found. Four of the five types of LDOCE warning notes were further investigated empirically to see whether they help correct errors and whether their visual salience influences success. In a naturalistic online survey, participants were given sentences to correct together with links to actual LDOCE entries, which featured the four types of warning notes, all of which included examples of errors. The results reveal widely varied error correction success rates ranging from 19% to 72%. Yet, the participants' internet activity was not in any way monitored to ascertain that the links were clicked and entries accessed. It is not known if the warnings were seen and read. While the study fails to provide evidence that the investigated types of warning notes affect error correction, it recognizes their potential in this respect, on condition that they are adequately conspicuous. Conspicuity, in turn, was observed to depend on the visual enhancement used (with borders and headings being the most effective) as well as prominent placement in the entry. As existing entries were employed, the place of warning notes was not controlled. It was nonetheless noted that warnings which were lower down the entries appeared less accessible (Nural, Nesi & Çakar 2022, p. 16). It is also interesting to point out that respondents considered warning types with more visual enhancements to be more useful. This preference did not tally with empirical results, which did not reveal any simple correspondence between the amount of highlighting in notes and error correction success. Although the study proves somewhat inconclusive, it is a pioneering one devoted entirely to warning notes featuring examples of errors. Its main contribution to the field consists probably in a typographical classification of warnings and the conclusion about the glaring inconsistency in the application of enhancement techniques (cf. Nural, Nesi & Çakar 2022, p. 16).

In a more controlled experiment, Dziemianko (2024) investigated the usefulness of examples of typical learner errors in MLDs for the accuracy of error correction, immediate and delayed retention of usage. Two types of purpose-built entries were employed. One included only regular examples, in the other examples of errors were added in red. No warning boxes were present. The position of examples of errors was manipulated to determine its effect on error correction and learning usage; they were placed in entry initial, medial and final positions. The results reveal that examples of errors failed to contribute significantly to error correction accuracy or immediate retention of usage, but largely enhanced delayed retention. In fact, they prevented retention of usage from deteriorating over time; thanks to them, delayed retention remained at a level similar to immediate retention. However, although their contribution to error correction accuracy was not *statistically significant*, it had undeniable *practical significance*; examples of errors helped to rectify about 50% more errors than regular examples. The positioning of examples

eISSN: 2550-2131
ISSN: 1675-8021

of errors in entries had no effect on error correction accuracy or the retention of usage. Overall, examples of errors were found to be a valuable standalone dictionary component, which may well be placed outside warning boxes.

None of the studies reviewed above compares the effectiveness of induction-oriented, standalone examples of errors with that of deductive warning boxes, which spell out rules of language and usually give examples of correct and incorrect usage. Arguably, the incorporation of examples of errors without warning boxes appears justifiable from the pedagogical perspective. Standalone examples of errors may prevent drawing false analogies and increase the accuracy of induction by showing which syntactic patterns are impossible. Research shows that explicit error correction by writing a learner sentence with an error on the board, crossing it out and writing a correct one above is the most pedagogically advantageous irrespective of the source of the error (L2 overgeneralization or transfer from L1, Tomasello & Herron, 1989). This implies that standalone examples of errors in dictionaries may likewise be beneficial for language learning. Tomasello and Herron (1989) also observe that preventing learners from committing errors by explicit warnings makes learning much less effective than correcting actually made learner errors. Possibly, then, placing examples of errors in dictionaries with their correction might be more pedagogically advantageous than citing explanations of linguistic rules without negative evidence.

By contrast, other studies suggest that direct explanations of rules and patterns in warning boxes may benefit language learners. Eye-tracking research shows that explicit metalinguistic explanations of target grammatical constructions increase attentional processing and successfully attract learners' cognitive resources to the described structures (Indrarathne & Kormos, 2017). Similarly, Varnosfadrani and Basturkmen (2009) found explicit feedback, including a metalinguistic explanation of the erroneous structure the most effective type of error correction. In computer-assisted language learning, Heift (2004) examined the role of corrective feedback in shaping learner uptake, i.e., learners' responses to the feedback. It turned out that providing an error explanation together with highlighting the error itself in the student input was the most effective form of corrective feedback, which generated the most learner uptake.

Implicit in direct error correction (whether with or without additional metalinguistic explanation) is the assumption that it helps learners to progress in L2. Error correction sensitizes learners to errors which may have become fossilized and are difficult to notice even at advanced levels, since they usually do not impede communication. That is why advanced adult learners in particular seem to need their errors made explicit to them to progress in developing L2 competence and prevent fossilization. Nonetheless, direct error correction is also claimed to do more harm than good and have no practical significance for learners' real long-term ability to use language communicatively in writing or speaking outside artificial grammar tests (e.g., Truscott, 2007).[5] Such an approach might justify omitting any examples of errors, standalone or otherwise, from learners' dictionaries.

The above review suggests that rationale can be found for including standalone examples of errors in MLDs, accompanying them with metalinguistic comments in warning boxes, as well as giving no warnings. The empirical study reported below tests the actual usefulness of all three solutions.

[5] Truscott (2007, p. 271) even argues that the question "*How effective is correction?* should be replaced by *How harmful is correction?*"

eISSN: 2550-2131
ISSN: 1675-8021

AIM

The paper aims to investigate the effect of warning messages in MLDs on error correction accuracy as well as immediate and delayed retention of usage. Two types of warnings are tested: standalone examples of errors placed outside any boxes (inductive), and warning boxes including usage explanations and examples of errors (deductive). The following research questions are posed:

RQ1. Is error correction accuracy affected by warnings in dictionary entries?
RQ2. Do warning messages in entries influence immediate and delayed retention of usage?
RQ3. Does the placement of warnings in entries affect error correction accuracy, immediate and delayed retention?

METHODS

MATERIALS

To answer the research questions, an online experiment was designed, which consisted of a pre-test, a main test, and immediate and delayed post-tests. All of them were based on 18 sentences with learner errors from the *Longman Dictionary of Common Errors* (Turton & Heaton, 1999; LDCE, see Table 1). In the pre- and posttests, participants corrected the sentences without dictionary support. The pre-test aimed to evaluate their lexical knowledge before the study and eliminate any cases where dictionary use was not necessary to correct the errors. The two post-tests checked the subjects' ability to correct the errors from memory immediately after exposure to dictionary information and two weeks later. In this way, immediate and delayed retention of correct usage was evaluated.

TABLE 1. Sentence cues used in the main test (with corrections)

Sentence cue in the main test	Correction	Keyword
The machine is supplied with instructions how to use it.	on how	instruction
'You are late!' she said with an angry voice.	in an angry voice	voice
Every day they voted what they would do the next day.	voted on	vote
Several passers-by stopped to look at the strange bike from curiosity.	out of curiosity	curiosity
All the prisoners had committed heavy crimes.	serious crimes	crime
Most of the damage has been produced by acid rain.	damage has been caused	damage
I hadn't made any experience of changing a car wheel.	hadn't had any experience	experience
The police are in favor of strict punishment.	severe punishment	punishment
Their services are very appreciated by the hotel management.	greatly appreciated	appreciated
He wanted to be with his son who was badly ill.	seriously ill	ill
What standards should we judge them with?	judge them by	judge
Have you heard what happened to the last patient he operated?	he operated on	operate
She made me so annoyed that I felt like shouting to her.	shouting at her	shout
The novel has been translated to English and French.	translated into	translate
We went to the party by a friend's car.	in a friend's car	car
The trade agreement will benefit for both parties.	benefit both parties	benefit
She spends most of her free time on reading.	spends time reading	spend
They spent the whole night fighting against the fire.	fighting the fire	fight

In the main test, error correction was based on purpose-built monolingual dictionary entries for the keywords involving errors in sentence cues (see Table 1). Each entry was based on the content of LDOCE and OALD, and consisted of a headword followed by pronunciation, a POS

eISSN: 2550-2131
ISSN: 1675-8021

label, a definition and six examples of usage, only one of which was relevant to the error correction task. The entries differed in the *placement* of useful syntactic information and *warnings*. The place of relevant examples was manipulated: they were at the top of the entry (in the first tercile), in the middle (in the second tercile) or at the bottom (in the third tericle). Three test versions were created, depending on the presence and type of warnings in entries:

1. with standalone examples of errors following useful examples,
2. with warning boxes following useful examples,
3. with no warnings (see the Appendix).

In the test versions where they were present, warnings (examples of errors and warning boxes) always followed regular examples relevant to the task at hand. Examples of errors were created on the basis of regular examples from MLDs, which were changed to represent the incorrect patterns from sentence cues. To illustrate, the OALD example of *instruction*: *The plant comes with full instructions on how to care for it*, was changed into ~~*The plant comes with full instructions how to care for it*~~ to reflect the error in the sentence offered for correction: *The machine is supplied with instructions on how to use it*. In the test, examples of errors were shown in ~~strikethrough font~~.

Warning boxes were compiled on the basis of LDCE. They included: the heading *Warning note*, an explanation of the correct pattern (in **bold**) introduced by *We say*, an indication of the incorrect pattern (in **bold**) after *NOT*, and two examples: a regular one in *italics* and a corresponding example of error in ~~*strikethrough italics*~~ following *NOT*. The latter showed the same error as the sentence cue. The warning notes were framed and printed against a light blue background (see Figure 8).

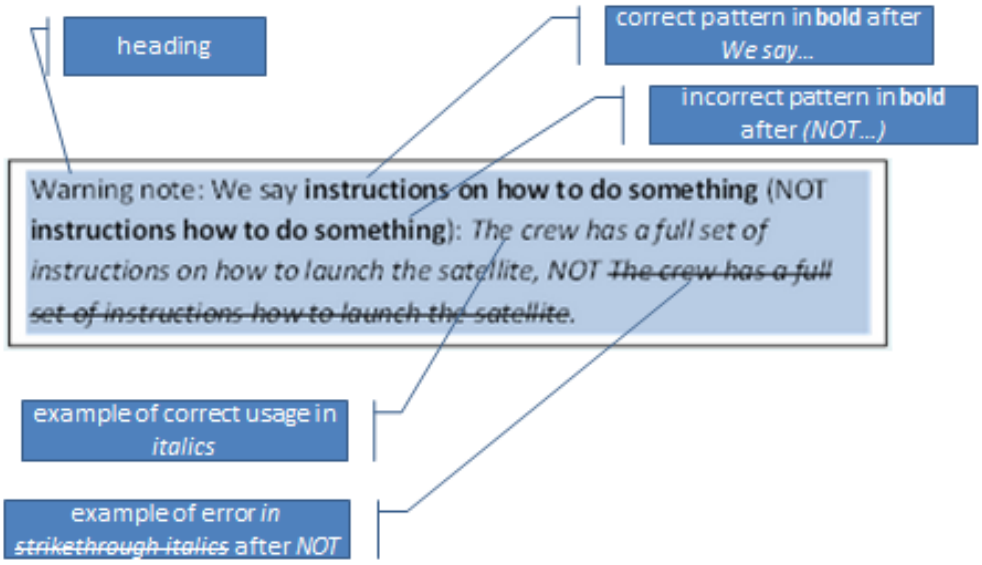


FIGURE 8. A warning box used in the experiment (with annotation)

PARTICIPANTS AND PROCEDURES

A total of 162 upper-intermediate (B2 in CEFR) Polish learners of English at Adam Mickiewicz University in Poznań participated in the study. Their level reflected that of the exam taken at the end of the academic year and the instruction materials done in class.

The experiment was conducted in a computer lab equipped with uniform PCs during standard classes. The participants were first requested to do the pre-test and correct the sentences shown to them without access to any sources. Immediately after the pre-test, they took the main test. A participant was randomly assigned to one test version; 55 students accessed the online dictionary with examples of errors, 52 consulted the version with warning boxes, and 55 were given the version with no warning messages. In the main test, the participants corrected errors in the sentences relying only on the supplied entries. Right after the main test, they took the immediate post-test, where the sentences had to be corrected from memory. The same post-test was repeated two weeks later.

At each stage of the study the sequence of the sentences with errors was randomized to reduce the learning effect. The experiment was self-paced inasmuch as there was no limit on the time needed to do individual tasks; the participants decided on their own how long they wanted to spend performing them. Yet, they could not go back and change any already given answer, but had to move sequentially through the questions. Throughout all the experiment, the participants' internet activity was closely monitored.

SCORING

For a student's answer to score a point, it had to concern the target error (rather than any other sentence part) and include its appropriate correction. To illustrate, no point was given when in the sentence *Several passers-by stopped to look at the strange bike from curiosity*, the collocation *from curiosity* was not corrected *into out of curiosity*, but another part of the sentence was (wrongly) considered incorrect, e,g., *passers-by* was rewritten as *passer-bys, by-passers* or *passerbies.* Similarly, no point was scored when the participants did identify the error, but failed to adequately correct it (e.g., because of a wrong choice of preposition: *for/by/at curiosity*). Also, the cases where instead of the target word participants used synonymous constructions or paraphrases were not awarded any points, e.g., *she said angrily* or *she said in an angry tone* instead of *she said in an angry voice*. Such answers did not result from reference to the supplied dictionary entry for *voice*, which was to be consulted to correct the error *she said with an angry voice*. On the other hand, spelling errors in the target words which did not affect meaning were ignored (e.g., *apreciated, appresiated, seriosly, operrated*).

To analyze the subjects' responses, two raters were involved. One of them was a native speaker of English, and the other was a native speaker of Polish proficient in English (C2 in CEFR). Both of them had over 20 years of experience teaching EFL at the academic level. The raters evaluated the participants' responses independently of each other. Any cases of divergent scores were discussed, and a consensus was reached. For example, one rater did not accept the sentence *Their service is greatly appreciated by the hotel management* as a correction of the original sentence *Their services are very appreciated by the hotel management* on account of the needless change in the number of the subject and the verb. In a discussion, the other rater conceded that the change was superfluous, but pointed out that the target structure *very appreciated* was nonetheless appropriately corrected for the adverbial (*greatly appreciated*). Upon reflection, the first rater admitted that the response deserved a point.

eISSN: 2550-2131
ISSN: 1675-8021

RESULTS

MAIN FINDINGS

There were two independent variables in the study: *warning messages* (with three levels: examples of errors, warning boxes, no warnings) and *target information position* in the entry (also with three levels: initial, medial, final).[6] The former was a between-groups factor, because each subject had access to only one dictionary version: with examples of errors, with warning boxes or without warning messages. The latter was a within-subject factor, because a participant consulted entries where target information was given in three places in equal measure (six entries with the information at the beginning of the entry, six different entries with the information in the middle, and the remaining six entries with the information at the end). The study investigates the influence of the factors on three dependent variables: error correction accuracy in the main test, immediate retention, and delayed retention. To analyze the data, a 3x3 between-within multivariate analysis of variance (MANOVA) was conducted. The MANOVA results reveal a statistically significant effect of *warnings* (Wilks's lambda=0.07, F=12.34, $p<0.001$, partial $\eta^2=0.593$), but not of *position* or their interaction ($p>0.05$). To further explore the significant effect, a univariate ANOVA was conducted for each dependent variable. Significant ANOVA results were further analyzed with the Bonferroni test.

The univariate ANOVAs indicate that *warnings* had a significant effect on error correction in the main test (F=18.68, Bonferroni corrected $p<0.001$, partial $\eta^2=0.713$), immediate retention (F=28.33, Bonferroni corrected $p<0.001$, partial $\eta^2=0.791$) and delayed retention (F=74.70, Bonferroni corrected $p<0.001$, partial $\eta^2=0.909$; see Figure 9). The results of post-hoc Bonferroni tests show that warning boxes were more useful for error correction, immediate and delayed retention of usage than examples of errors and entries without warnings ($p<0.05$). Examples of errors helped in all these respects more than no warnings in entries ($p<0.05$).

It is particularly noteworthy that delayed retention virtually slumped in the absence of warnings. Participants consulting examples of errors (53.21%) and warning boxes (66.16%) remembered in the long run, respectively, about two and a half times and three times more than learners not exposed to any warnings in entries (21.52%; 53.21*100/21.52=247.26; 66.16*100/21.52=307.43).

TABLE 2. Results of post-hoc Bonferroni tests (*p*-values) by dependent variable

Dependent variables:	**Error correction in the main test**			**Immediate retention**			**Delayed retention**		
Correction accuracy (%)	**(1) 56.01**	**(2) 67.97**	**(3) 79.91**	**(1) 42.78**	**(2) 62.19**	**(3) 75.08**	**(1) 21.52**	**(2) 53.21**	**(3) 66.16**
(1) no warnings	x	0.02	0.00	x	0.00	0.00	x	0.00	0.00
(2) examples of errors	0.02	x	0.02	0.00	x	0.03	0.00	x	0.01
(3) warning boxes	0.00	0.02	x	0.00	0.03	x	0.00	0.01	x

[6] Target information refers to relevant regular examples followed by warnings (where applicable).

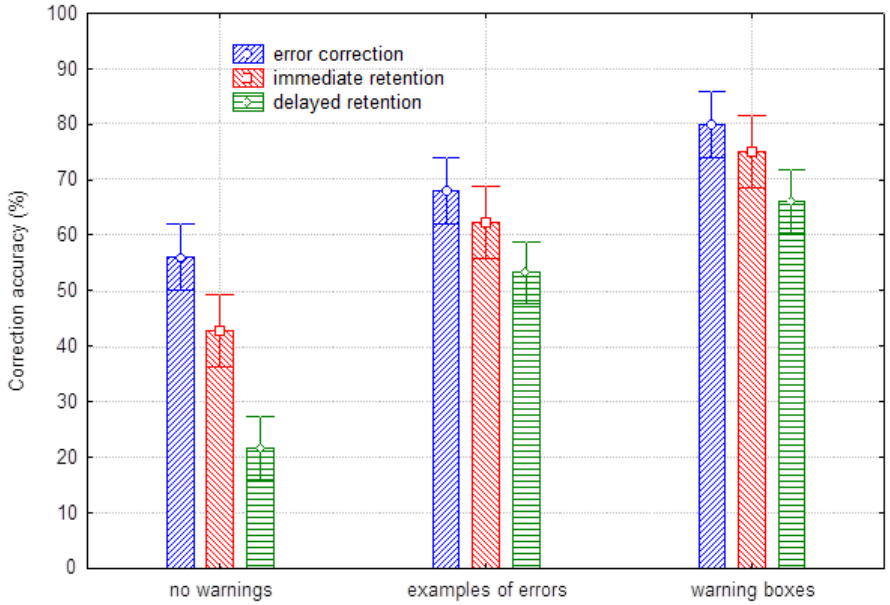


FIGURE 9. Error correction accuracy, immediate and delayed retention of usage by warning message in entries. Vertical bars denote 95% confidence intervals

ANCILLARY FINDINGS

The analysis of error correction accuracy for each sentence cue in the main test reveals a consistent advantage of warning boxes over the other two experimental conditions (see Figure 10). First, warning boxes were always more helpful than no warnings. For *voice* and *punishment*, the scores obtained with their help were as much as about 70% better than those based on entries without warnings (*voice*: 78.2*100/46.3=168.9; *punishment*: 79.3*100/47.5=167.0). In five cases (*instruction, damage, shout, translate, fight*), the difference in favor of warning boxes exceeded 50%. In the case of *spend*, on the other hand, users of warning boxes outstripped those consulting entries devoid of warnings only by 16% (75.2*100/64.9=115.9).

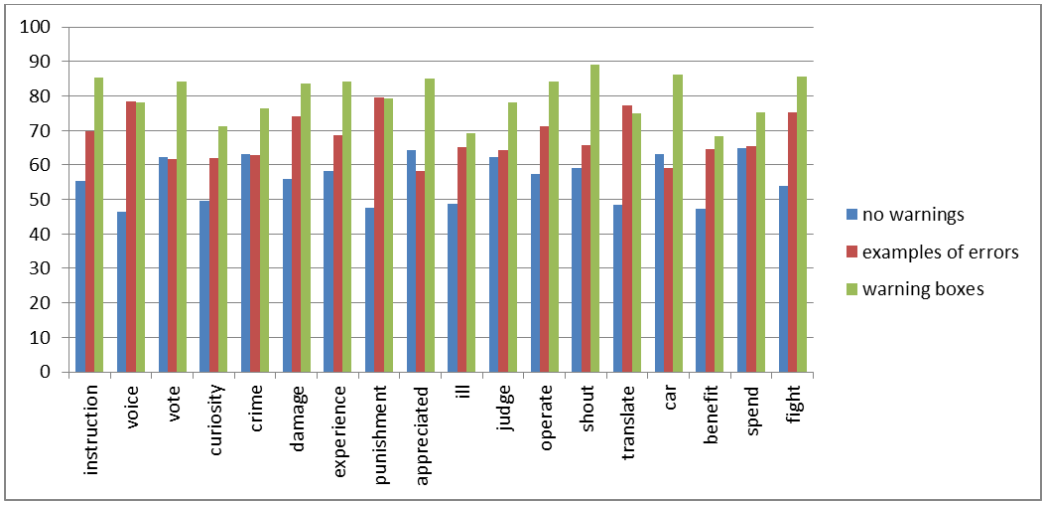


FIGURE 10. Error correction accuracy in the main test by error and warning in entries

Second, warning boxes were, in the vast majority of cases (83.3%), more helpful than examples of errors. Their greatest advantage was noted for *appreciated* and *car*, where they helped to rectify about half more errors than examples of incorrect usage (*appreciated*: 85.1*100/58.3=146.0; *car*: 86.3*100/59.1=146.0). It was the smallest for *ill* and *benefit*, where they improved error correction scores by about 6% (*ill*: 69.2*100/65.2=106.1; *benefit*: 68.4*100/64.6=105.9). Only in three cases (*voice, punishment, translate*) were examples of errors marginally more useful than warning boxes, their slight edge ranging from 0.4% for *voice* (78.5*100/78.2=0.4) and *punishment* (79.6*100/79.3=100.4) to around 3% for *translate* (77.2*100/75.1=102.8).

Third, in over three-fourths of all cases (77.8%), examples of incorrect usage helped to correct errors more successfully than entries with no warnings. They improved error correction accuracy by as much as 70% in the case of *voice* and *punishment* (*voice*: 78.5*100/46.3=169.5, *punishment*: 79.6*100/47.5=167.6) and almost 60% in the case of *translate* (77.2*100/48.4=159.5). Yet, for *spend* and *judge*, their contribution to error correction success was marginal (*spend*: 65.4*100/64.9=100.8, *judge*: 64.3*100/62.4=103.0). In four cases (*crime*, *vote*, *car* and *appreciated*), the results based on entries without warnings were better than those grounded on entries featuring examples of errors, with the advantage of the former ranging from 0.3% for *crime* (63.1*100/62.9=100.3) to over 10% for *appreciated* (68.4*100/58.3=110.3).

SUMMARY OF THE RESULTS

The study gives affirmative answers to the first two research questions: warnings in MLD entries influence error correction accuracy (RQ1), immediate and delayed retention (RQ2). In all three respects, boxes are the most beneficial, followed by examples or errors and no warnings. The third research question needs to be answered negatively; the place of information useful for error correction (warnings and relevant regular examples) in MLD entries does not influence error correction accuracy and retention (RQ3). It thus turns out that warning boxes are the most recommendable type of warning messages in entries, because they help most to correct errors and learn usage. Examples of errors are second best, while entries without warnings are the least helpful. Overall, the research shows that warnings are a perfect complement to the dictionary microstructure.

DISCUSSION

The results, which reveal that warning boxes are more useful than examples of errors coincide with previous findings pointing to the greater effectiveness of the deductive approach to teaching grammar (represented by warning boxes) than the inductive one (embodied by standalone examples of errors; cf. Shirav & Nagai, 2022). The study suggests that examples of errors are also a valuable contribution to the dictionary microstructure, even though they are not as efficient as warning boxes. Still, they substantially improve error correction accuracy, immediate and delayed retention in comparison with no warnings. On the one hand, such findings confirm those drawn by Dziemianko (2024), where examples of errors were also found a welcome supplement to MLD entries. On the other, they diverge from those in the (2024) study inasmuch as examples of errors had there no statistically significant effect on error correction accuracy. The divergence might be due to the different visual enhancement of examples of errors in both studies (~~strikethrough~~ in the current experiment vs red in the previous one). The former highlighting technique leaves no doubt

that the text is wrong. The latter conventionally performs a warning function, implies caution, importance or danger, but does not have to mean that the text is incorrect; it may well be important (cf. Strobelt et al., 2016, p. 492). Possibly, the participants directly associated strikethrough examples with incorrectness and successfully used them in error correction.

Chan (2012) found that warning boxes were largely neglected by learners. In the current study, by contrast, they were successfully consulted, possibly due to the experimental conditions. In this investigation, the presence, position, content and highlighting of warning boxes were systematically controlled in purpose-built entries, while in the study by Chan (2012) they were not, as entries copied from CALD3 were employed. In particular, "Common mistake" boxes might have been ignored because they usually come at the end of CALD3 entries or subentries (if more than one part of speech are lumped together). Having already found the needed information earlier in the entry, the participants might not have been bothered to continue reading. In the current study, warning boxes were evenly distributed in entry initial, medial and final positions. Maybe for this reason they did not seem so negligible. As for visual enhancement, in both investigations warning notes involved background coloring (light blue in the current study vs light turquoise in the other one). However, in the present experiment, they were additionally framed. In CALD3, they were not. According to Strobelt et al. (2016), surrounding a text by a border makes it stand out; it appears bigger and is more easily detectable. That might also be a reason why boxed warnings were more useful. Another might be the Hawthorne effect; the participants assigned to the test version with warning boxes might have felt expected to consult them. It is not certain, however, that they would benefit from them in more naturalistic contexts.

LIMITATIONS

The efficacy of induction and deduction may depend on learners' preferences, learning styles or language aptitude. Such factors were not considered in the current analysis. Besides, error categories were not controlled, so it is not known if they affected the usefulness of warnings. In some sentence cues, prepositions were incorrect (e.g., *with an angry voice*), superfluous (e.g., *to benefit for both parties*) or left out (e.g., *instruction how to*). Some other errors consisted in wrong adjectives in adj+N collocations (e.g., *heavy crimes*), verbs in v+N collocations (*produce damage*) or adverbs in adv+ADJ collocations (*very appreciated*). Interestingly, Tono et al. (2014) show that some errors (omission and addition) are more suitable for checking against corpus data than others (misformation); corpus consultation brings higher correction accuracy rates for the former than for the latter. Warnings based on learner corpora might likewise be more or less useful for some error types. Also, the effectiveness of warnings may depend on the degree to which learner errors are motivated by negative transfer from L1. If some errors recur among speakers from a given native language background due to incongruities between English and their L1, explicit warnings against such mistakes may be more needed than against others. The current study did not address this issue.

FURTHER RESEARCH AND RECOMMENDATIONS

While warnings in MLDs prove to be highly beneficial to error correction, it remains to be seen if they could help learners to *avoid* committing errors in natural tasks like writing essays or research papers. To answer such a question, the incidence of errors in assignments written with the help of dictionary warnings and without it by learners representing the same proficiency could be

compared. To make their output sufficiently comparable, a list of error-prone words to be included in writing could be supplied. Also, as the participants in the current study were all native speakers of Polish, it is not known if similar results could be obtained for speakers of other languages, whose different L1s may make other aspects of English grammar challenging.

Research is needed to see how various formatting conventions of highlighting text affect the perception of warnings and learning usage. Interestingly, a series of crowdsourced experiments showed that the effectiveness of common web-friendly text highlighting techniques depends on whether they are used individually or jointly (Strobelt et al., 2016). For example, while increasing font size is the most effective single technique, significant visual interferences occur when it is combined with other methods (e.g., border). Background coloring and text coloring, in turn, rarely interfere with other techniques. Such caveats should be considered when the highlighting of warnings in MLDs is decided. Ideally, the effectiveness of different (constellations of) visual enhancements should be tested empirically among learners so that their application to dictionary warnings would no longer be arbitrary or discretionary.

It is also necessary to investigate what exactly to highlight in warning messages. For example, Lavie et al. (2004) found that visual enhancement of text sections increases perceptual load and may result in the exclusion of the stimulus from further cognitive processing unless the stimulus is deemed relevant to the task. As warnings in MLDs are not adjusted to the needs of speakers of any L1, but address those of learners worldwide, it might be desirable to increase their relevance and introduce L1-specific customization into the dictionary interface. Such customization should make it possible to call up the errors which result from interference from a given L1 and are more suitable for its speakers.

Interestingly, attempts at customizing dictionaries to mother tongue backgrounds have already been made, for example in the Louvain EAP Dictionary (LEAD), a web-based English for Academic Purposes dictionary-cum-writing aid for non-native writers. Before the look-up, it requests the user to select the mother tongue (*French, Dutch, Spanish, Chinese, German* and *Other*). L1-background identification enables giving contrastive feedback on errors that native speakers of a given L1 typically commit (Paquot, 2012, p. 175, 178). The dictionary offers two types of warnings: generic and L1-specific. Errors which occur in a wide range of learner populations are treated in generic error notes displayed irrespective of the selected mother tongue (cf. Figure 11, which shows that learners tend to use *it* as a subject after *as*). Those which are L1-specific, in turn, appear in notes which show up when an L1 is chosen and highlight major differences between English academic words and their error-prone translation equivalents, e.g., the erroneous translation of *selon moi* into *according to me* by French users (Figure 12, cf. Paquot, 2012, p. 179). It seems that warnings in MLDs would largely benefit from a similar degree of customization.

eISSN: 2550-2131
ISSN: 1675-8021

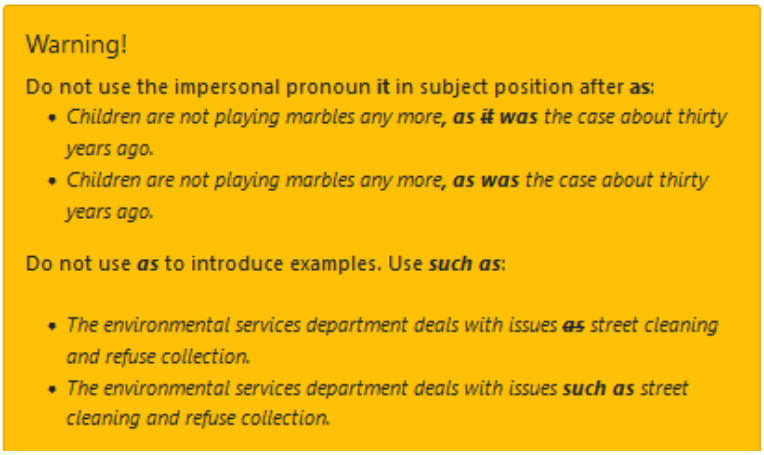


FIGURE 11. Warnings in the entry for *as* in LEAD

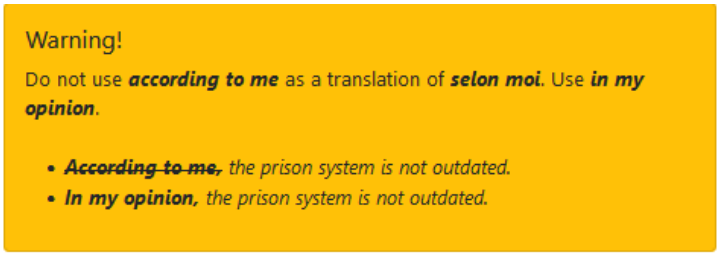


FIGURE 12. Warnings in the entry for *according to* in LEAD

CONCLUSIONS

The study investigates the role of examples of errors and descriptive warning messages in MLDs. It provides empirical evidence for the usefulness of both warning types in online learners' dictionaries for error correction as well as immediate and delayed retention of usage. It also reveals that deduction-oriented warning boxes, which explain and illustrate correct and incorrect language use, are more advantageous in these respects than induction-oriented standalone examples of errors, which provide no metalinguistic comment. The latter, in turn, are more helpful for error correction and learning than no warnings. It thus appears that in the AI era, when dictionaries are being superseded by interactive generative AI applications, warnings can be a compelling feature of MLDs. Interestingly enough, they seem to remain the hallmark of conventional reference tools which has not been thoroughly investigated yet in the context of AI. Recent publications about the usefulness of AI for typical lexicographic tasks do not concern generating dictionary warnings, either in the form of examples of errors or warning messages. The AI-generated entry components arousing researchers' interest typically involve definitions, examples of correct usage, word forms and word senses, pronunciation, collocations, synonyms or antonyms (e.g., Jakubicek & Rundell, 2023; Rundell, 2023; Lew 2023). While the involvement of large language models in lexicographers' benchmarks has generally engendered enthusiasm, opinion is still divided as to how far lexicographers can be superseded by AI (cf. De Schryver, 2023, Rundell, 2023). It appears that alerting learners to incorrect language usage in dictionaries remains (for now) the domain of humans rather than machines.

eISSN: 2550-2131
ISSN: 1675-8021

REFERENCES

Béjoint, H. (1994). *Tradition and Innovation in Modern English Dictionaries*. Oxford: Clarendon Press.

Cambridge advanced learner's dictionary. Retrieved December 5, 2024 from https://dictionary.cambridge.org/pl/dictionary/english

Chan Yin-Wa Alice. (2012). Cantonese ESL Learners' Use of Grammatical Information in a Monolingual Dictionary for Determining the Correct Use of a Target Word. *International Journal of Lexicography*. *25*(1), 68-94. https://doi.org/10.1093/ijl/ecr014

Chen, Y. (2017). Dictionary use for collocation production and retention: A CALL-based study. *International Journal of Lexicography*. *30*(2), 225-251. https://doi.org/10.1093/ijl/ecw005

Collins Online Dictionary. Retrieved November 25, 2024 from https://www.collinsdictionary.com/dictionary/english

De Cock, S. & Granger, S. (2004). Computer learner corpora and monolingual learners' dictionaries: The perfect match. *Lexicographica*. *20*, 72-86. https://doi.org/10.1515/9783484604674.72

De Schryver, G.-M. (2023). Generative AI and lexicography: The current state of the art using ChatGPT. *International Journal of Lexicography*, *36*(4), 355-387. https://doi.org/10.1093/ijl/ecad021

Dziemianko, A. (2006). *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English*. Max Niemeyer Verlag.

Dziemianko, A. (2006). *User-friendliness of verb syntax in pedagogical dictionaries of English*. Max Niemeyer Verlag.

Dziemianko, A. (2014). On the presentation and placement of collocations in monolingual English learners' dictionaries: Insights into encoding and retention. *International Journal of Lexicography*, *27*(3): 259-279.

Farina, D., Vrbinc M. & Vrbinc A. (2019). Problems in online dictionary use for advanced Slovenian learners of English. *International Journal of Lexicography*. *32*(4), 458-479. https://doi.org/10.1093/ijl/ecz017

Frankenberg-Garcia, A. (2012). Learners' use of corpus examples. *International Journal of Lexicography*. *25*(3), 273-296. https://doi.org/10.1093/ijl/ecs011

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL*. *26*(2), 128-146. https://doi.org/10.1017/S0958344014000093

Frankenberg-Garcia, A. (2015). Dictionaries and encoding examples to support language production. *International Journal of Lexicography*. *28*(4), 490-512. https://doi.org/10.1093/ijl/ecv013

Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL*. *16*(2), 416-431.

Hornby, A. S. (1956). *A Guide to Patterns and Usage in English*. London: Oxford. University Press.

Indrarathne, B. & Kormos, J. (2017). Attentional processing of input in explicit and implicit learning conditions: An eye-tracking study. *Studies in Second Language Acquisition*. *39*(3), 401-430.

Jakubíček, M., & Rundell, M. (2023). The end of lexicography? Can ChatGPT outperform current tools for post-editing lexicography? In M. Medveď, M. Měchura, I. Kosem, J. Kallas, C. Tiberius, & M. Jakubíček (Eds.), *Proceedings of the eLex 2023 Conference* (pp. 518-533). Brno: Lexical Computing.

Lavie, N., Hirst, A., De Fockert, J. W. & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*. *133*, 339-354.

Lew, R. (2023). ChatGPT as a COBUILD lexicographer. *Humanities and Social Sciences Communications*, *10*(1), 704. https://doi.org/10.1057/s41599-023-02119-6

Longman Dictionary of Contemporary English Online. Retrieved December 5, 2024 from http://www.ldoceonline.com/

eISSN: 2550-2131
ISSN: 1675-8021

Macmillan Dictionary English. Retrieved November 25, 2024 from https://play.google.com/store/apps/details?id=dictadv.english.britishmacmillan

Nesi, H. & Tan, K. H. (2011). The effect of menus and signposting on the speed and accuracy of sense selection. *International Journal of Lexicography*. *24*(1), 79-96. https://doi.org/10.1093/ijl/ecq040

Nural, S., Nesi, H. & Çakar, T. (2022). Warning notes in a learners' dictionary: A study of the effectiveness of different formats. *International Journal of Lexicography*. *35*(4), 449-467. https://doi.org/10.1093/ijl/ecab033

Osada, T., Sugimoto, J., Asada, Y. & Komuro, Y. (2015). An analysis of *Longman Dictionary of Contemporary English*, sixth edition. *Lexicon. 45*, 1-73.

Oxford Advanced Learner's Dictionary. Retrieved December 5, 2024 from https://www.oxfordlearnersdictionaries.com/

Paquot, M. (2012). The LEAD dictionary-cum-writing aid: An integrated dictionary and corpus tool. In S. Granger & M. Paquot, (Eds.), *Electronic Lexicography*. Oxford: Oxford University Press.

Ptasznik, B. (2024). A wealth of information or too much information? Examining the effectiveness of supplementary corpus examples in online dictionaries. *GEMA® Online Journal of Language Studies*. *24*(1), 18-37. http://doi.org/10.17576/gema-2024-2401-02

Rundell, M. (2023). Automating the creation of dictionaries: Are we nearly there? In Y. An (Ed.), *Proceedings of the 16th International Conference of the Asian Association for Lexicography* (pp. 1-9). Seoul: Yonsei University.

Shirav, A. & Nagai, E. (2022). The effects of deductive and inductive grammar instructions in communicative teaching. *English Language Teaching*. *15*(6), 102-123.

Strobelt, H., Oelke, D., Kwon, B. C., Schreck, T. & Pfister, H. (2016). Guidelines for effective usage of text highlighting techniques. *IEEE Transactions on Visualization and Computer Graphics*. *22*(1), 489-498. https://doi.org/10.1109/TVCG.2015.2467759

The Louvain EAP Dictionary. Retrieved November 25, 2024 from https://leaddico.uclouvain.be

Tomasello, M. & Herron, C. (1989). Feedback for language transfer errors: The garden path technique. *Studies in Second Language Acquisition*. *11*(4): 385-395. https://doi.org/10.1017/S0272263100008408

Tono, Y., Satake, Y., & Miura, A. (2014). The effects of using corpora on revision tasks in L2 writing with coded error feedback. *ReCALL*. *26*(2), 147-162. https://doi.org/10.1017/S095834401400007X

Truscott, J. (2007). The effect of error correction on learners′ ability to write accurately. *Journal of Second Language Writing. 16*, 255-272.

Tsai, K.-J. (2019). Corpora and dictionaries as learning aids: Inductive versus deductive approaches to constructing vocabulary knowledge. *Computer Assisted Language Learning*. *32*(8), 805-826. https://doi.org/10.1080/09588221.2018.1527366

Turton, N. & Heaton, J. (1999). *Longman Dictionary of Common Errors*. London: Longman.

Varnosfadrani, A. D. & Basturkmen, H. (2009). The effectiveness of implicit and explicit error correction on learners' performance. *System*. *37*, 82-98.

APPENDIX

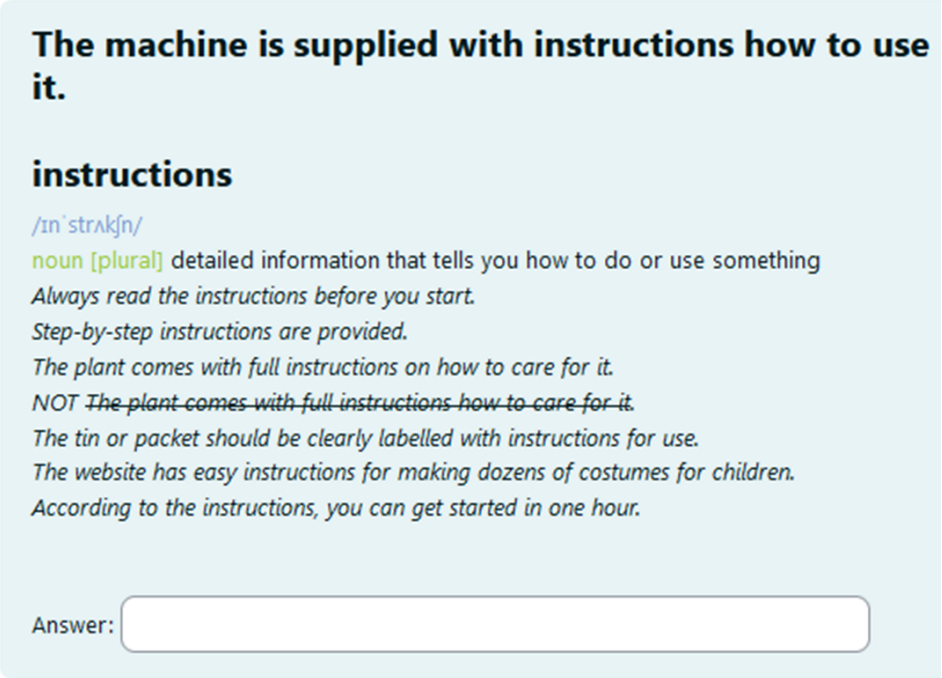


FIGURE 5. Test version with examples of errors

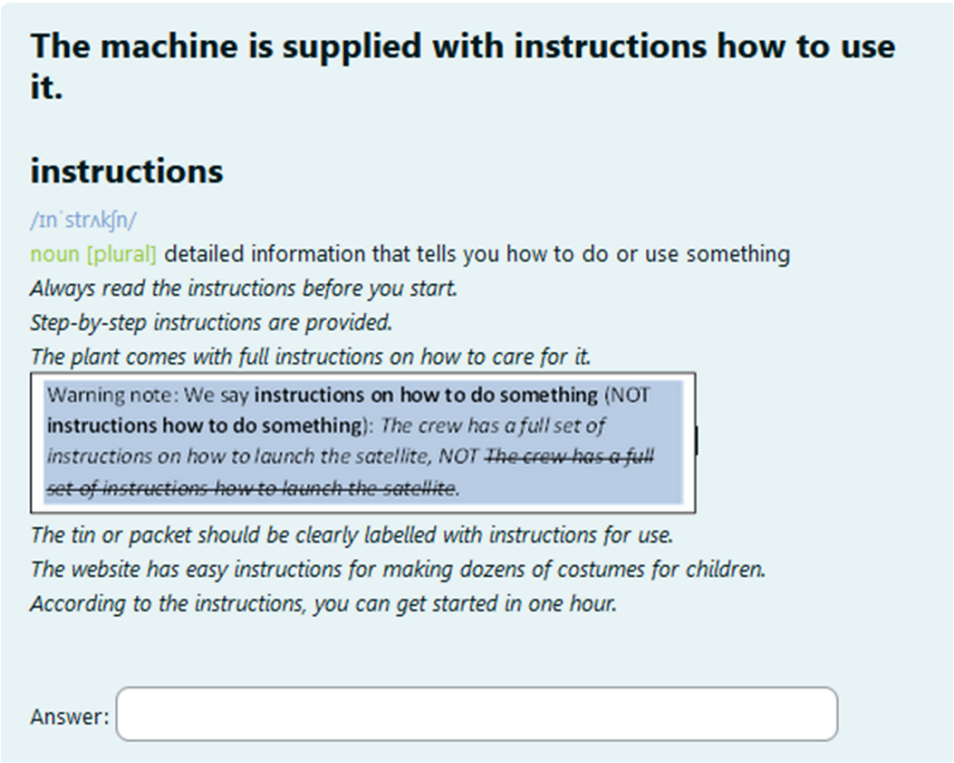


FIGURE 6. Test version with a warning box

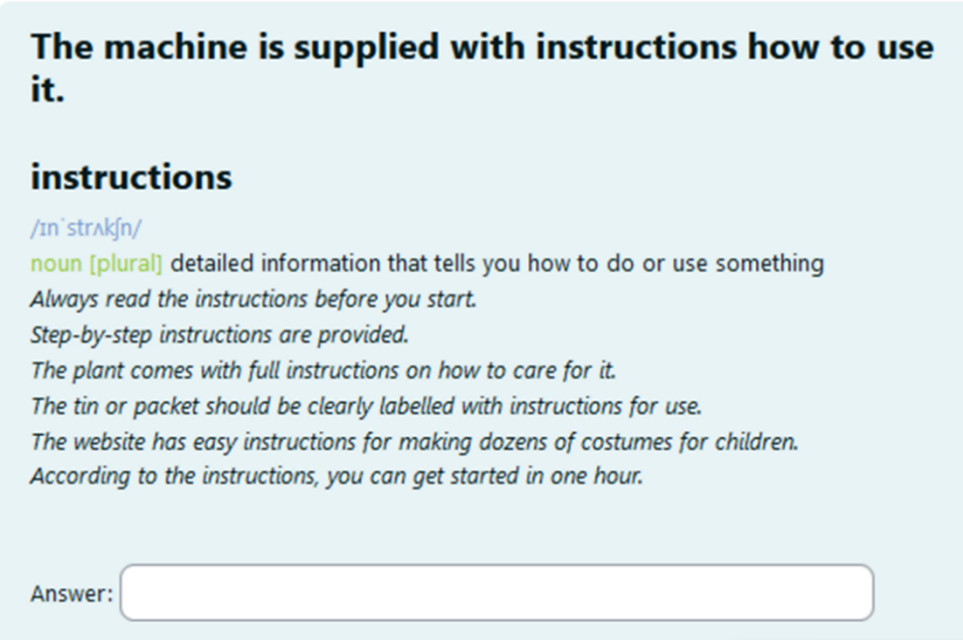


FIGURE 7. Test version with no warnings

ABOUT THE AUTHOR

Anna Dziemianko is a university professor at Adam Mickiewicz University in Poznań, Poland. Her research interests include English lexicography, user studies, syntax in dictionaries, defining strategies and the presentation of lexicographic data on dictionary websites. Recently she has been conducting comparative user research into AI, web search and dictionaries.

eISSN: 2550-2131
ISSN: 1675-8021