ReCALL (2024), 36: 1, 89–103
doi:10.1017/S0958344023000241

RESEARCH ARTICLE

Do we learn from mistakes? The usefulness of examples of errors in online dictionaries

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

Abstract
The aim of this paper is to investigate the usefulness of examples that show typical learner errors in online pedagogical dictionaries of English for the accuracy of error correction as well as immediate and delayed retention of usage. The optimal positioning of examples of errors in entries is also researched. In an online experiment, participants did a sentence correction exercise with the help of purpose-built monolingual dictionary entries, where the provision and positioning of examples showing errors were controlled. Two test versions were created, which differed only in the presence of examples of errors in the entries. Usage retention was observed immediately after the test and two weeks later. The results indicate that it is worthwhile to include examples of errors in online learners' dictionaries because they contribute greatly to the retention of usage in the long run. They also help to rectify errors, though the effect is not statistically significant. The positioning of examples showing errors in entries has no influence on error correction accuracy or usage retention. The study reveals examples of errors to be a valuable induction-oriented stand-alone dictionary component placed outside warning boxes, which typically include explicit grammar rules and promote deduction.

Keywords: online dictionaries; learner errors; error correction; examples; retention

1. Introduction

Examples in monolingual English learners' dictionaries (henceforth, MLDs) show how English words are typically used. Learners may wonder, however, whether only the exemplified patterns are possible. After all, other usage might also be correct, though not common enough to deserve a place in a dictionary. Today, all prestigious MLDs are based on extensive corpus analysis, which, among others, informs the selection of examples. Yet, the curated 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: 287). This shortcoming seems to be partly overcome by examples of errors, which illustrate unacceptable usage. They are based on learner corpora, which "enable lexicographers to identify recurrent sources of difficulty, and ... use this information to anticipate learners' errors" (Rundell, 1999: 47). In contrast to regular examples, like *The film was absolutely hilarious*, examples of errors show how a language should not be used and warn against frequent linguistic problems (e.g. ✗*Don't say: The film was very hilarious*).[1]

Cite this article: Dziemianko, A. (2024). Do we learn from mistakes? The usefulness of examples of errors in online dictionaries. *ReCALL* 36(1): 89–103. https://doi.org/10.1017/S0958344023000241

[1]Examples of correct usage are referred to below as regular examples, while those that illustrate typical errors are called examples of errors.

© The Author(s), 2024. Published by Cambridge University Press on behalf of European Association for Computer-Assisted Language Learning. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

De Cock and Granger (2004) note that at the beginning of the 21st century, based on learner corpus analysis, the *Longman Dictionary of Contemporary English* (fourth edition; LDOCE4, 2003) and the *Cambridge Advanced Learner's Dictionary* (first edition; CALD1, 2003) included 250 "Warning Notes" and 200 "Common Learner Error" sections, respectively, to draw learners' attention to common linguistic failures (e.g. *I was surprised by *these news*, or *we discussed *about the plans for the wedding*). Apart from examples of errors, the notes spelled out rules of English grammar (e.g. "**News** is used with the singular form of words such as **this** and **that**," or "Remember that **discuss** is never followed by 'about'").

Warning sections are presumed to be particularly useful in MLDs. At advanced levels, many errors have become fossilized and remain unnoticed by learners because they do not hinder communication. Explicit negative feedback helps learners pay attention to persistent errors, which is a prerequisite for their eradication (Granger, 2008: 270). Yet errors are not always given; some warning boxes contain only grammar rules or explanations.[2] It is, however, highlighting examples of errors that prove to be particularly useful for language learning. Tomasello and Herron (1988, 1989) show that explicit error correction (writing an incorrect sentence on the board, crossing it out, and putting a correct one above) brings a considerable and lasting pedagogical advantage, regardless of whether learner errors result from second language (L2) generalization or first language (L1) transfer.

Explicit warnings against recurrent errors have their drawbacks, too. MLDs are targeted at users around the world and rely on learner corpus data from a range of L1s. Thus, they address cross-linguistic difficulties common to learners from different mother tongue backgrounds rather than L1-specific ones. Nonetheless, De Cock and Granger (2004: 82) notice that there is surprisingly little overlap across MLDs in their selection of headwords for which explicit error notes are supplied, international learner corpus foundations notwithstanding. Besides, the errors covered are often not frequent enough or inappropriate for the advanced proficiency level.

An important problem related to negative evidence concerns its placement in a dictionary. De Cock and Granger (2004: 82) note that when a few lexical items with separate entries are confused, such as *a great *amount* (instead of *number*) *of people*, it needs to be decided under which headword the relevant warning should be placed (*amount*, *number* or *people*). The positioning of a warning message seems to be less problematic in e-dictionaries, where hyperlinks ensure access from all the entries involved. Yet, recent research into dictionary users' behavior shows that hyperlinks are often ignored (Chen, 2017, 2020; Li & Xu, 2015). Placement problems appear even when warnings concern a headword treated in one entry. Entry parts do not have the same salience, but, oddly enough, it is not known which are the most salient (and where warnings could be optimally located). Although the beginning of the entry is usually considered the most prominent (e.g. Chen, 2017, 2020; Lew, Grzelak & Leszkowicz, 2013), there are experiments that point to the salience of entry bottom (Dziemianko, 2014, 2017; Nesi & Tan, 2011).

Another issue connected with warning notes concerns their visual presentation, which should ensure prominence and noticeability, but often does not (De Cock & Granger, 2004: 82). If adequately conspicuous, examples of errors can make learners pay attention to errors, whose conscious noticing, in keeping with the noticing hypothesis (Schmidt, 1990), may assist in language learning. The hypothesis assumes that if input is to be processed for L2 acquisition, it must first be noticed. Techniques of visual input enhancement (VIE) with typographic cues (e.g. **bolding**, CAPITALIZATION, *italicization*, underlining, color coding) make target forms perceptually salient and noticeable, facilitate processing, and are assumed to affect learning (Jourdenais, Ota, Stauffer, Boyson & Doughty, 1995; Lee & Huang, 2008). Yet, research into the role of VIE in grammar learning produces mixed results. Some studies indicate positive effects (e.g. Jourdenais *et al.*, 1995; Lee, 2007; Park, Choi & Lee, 2012); others do not (e.g. Lee & Jung, 2021; Winke, 2013; Wong, 2003). Even replicated research yields conflicting findings.

[2]To illustrate, examples of errors are absent from 20% of such boxes in LDOCE6 (Osada, Sugimoto, Asada & Komuro, 2015: 43).

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

Lee (2007) demonstrated that incorrect English passive forms were more successfully corrected by learners in the VIE condition than in the unenhanced condition. In a conceptual replication of Lee's (2007) experiment using eye-movement data, Winke (2013: 323–324) concluded that VIE, though noticed, did not affect error correction: "enhancement in this study … promoted noticing, but … appeared to have done little else." More recently, Loewen and Inceoglu (2016) observed in their eye-tracker study that visually enhanced grammatical forms do not stimulate learning, and do not attract attention, either.

Such contradictory results might be explained by widely different methodologies, including various typographical cues in the tasks. Simard (2009) shows that not all input enhancement exerts the same influence on learners; the effect differs according to the choice and number of typographical cues. She found that capital letters and three cues combined (capitalizing, underlining and bolding) entail better grammar test scores than seven other enhancement methods.

Rarely has VIE been researched in lexicography. Dziemianko (2015) noted that functional labels in color help to extract syntactic information from MLDs and improve its retention. Dziemianko (2014) observed that the visual presentation of collocations in MLDs affects encoding and retention. Lew *et al.* (2013) found that bold print firmly focuses dictionary users' gaze.

Overall, the inclusion of examples showing errors in MLDs, though potentially beneficial for users, entails difficult decisions on their placement in entries and highlighting. The following section reveals how these problems are currently resolved in lexicographic practice.

2. Examples of errors in MLDs

Examples of errors feature in major online MLDs: LDOCE, CALD, the *Macmillan Dictionary Online* (MDO), and the *Oxford Advanced Learner's Dictionary* (OALD). To get an insight into their presentation, Table 1 in the supplementary material collates relevant extracts.[3]

In OALD, examples of errors are given in pink-shaded warning boxes at the end of a relevant sense or immediately below an error-prone pattern. A warning box explains how a word can and/or cannot be used (e.g. You cannot 'suggest somebody something'), and includes examples of correct and incorrect usage, the latter in italicized strikeout font (e.g. ~~*Can you suggest me a better way of doing it?*~~, ~~*The property comprises of bedroom, bathroom and kitchen*~~). Normally, examples of errors take the form of full sentences, but incorrect patterns are occasionally illustrated without any sentential context – just put in parentheses in inverted commas after *not*, e.g. (not 'an advice'), (not 'some advices').

In LDOCE, examples of errors appear in three places. First, they are given in grammar or usage boxes, where correct usage is explained and exemplified. Such boxes follow a relevant sense or entry. Examples of errors are typically typed in red, preceded by a red cross (✗) and a "Don't say" note (✗ Don't say: The exam is comprising four questions). Not always are full-sentence examples given; only incorrect patterns may be spelled out (e.g. ✗ Don't say: by an accident, ✗Don't say: accuse someone to do something, ✗ Don't say: an advice).[4] Second, similar warnings can be found in collocation boxes under the "COMMON ERRORS" subheading. Then, a red triangle (►) and "Don't say" precede the incorrect collocation printed in canonical form and strikeout font. A "Say" note follows, along with a correct collocation in italics (►Don't say '~~make an experiment~~'. Say *carry out an experiment* or *do an experiment*). Third, some examples of errors are placed in LDOCE immediately after a relevant sense, outside any box. Then, the incorrect pattern appears in inverted commas in regular font, and is preceded by a red triangle and a "Don't say" note. It is followed by an example of appropriate usage in italics and a full-sentence example of errors in

[3]MDO closed down on 30 June 2023, when the first draft of the manuscript had already been written.

[4]The quotes faithfully reflect the inconsistent use of red.

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

parentheses after the capitalized word NOT: ►Don't say 'give to someone something': *He gave me a card (NOT He gave to me a card).*

In MDO, examples of errors appear in GET IT RIGHT! boxes at the end of the entry, where correct and incorrect usage is explained and exemplified. Correct usage is highlighted in bold in italicized regular examples flagged with a tick ✓(e.g. ✓ *The government* ***suggested constructing*** *another railway link to the mainland*). Incorrect usage is shown in bold in examples flagged with a cross ✗ (e.g. ✗ *The government* ***suggested to construct*** *another railway link to the mainland*). Interestingly, to discuss "common misconceptions about certain vocabulary questions," MDO features 10 short video clips titled Real Vocabulary. The one for *comprise* features a full-sentence example showing an error, and is flagged with a red cross. The error is highlighted in red on the screen as well as commented on in the clip.

CALD online offers examples of errors in clickable grammar notes whose heading is printed in bold navy-blue font on a yellow background. Upon clicking, the note unfolds and grammar rules are spelled out under relevant subheadings. Italicized examples of correct usage with either the headword or the whole patterns in bold follow respective rules (e.g. *My teacher* ***suggested*** *an exam* ***to me*** *which I could take at the end of the year*). Examples of errors, if present, are given below in strikeout font and are preceded by "Not:" (Not: ~~My teacher suggested me an exam~~). Some of them do not take the form of whole sentences (e.g. Not: . . . ~~a lot of advices~~). The last section in a CALD grammar note is titled "Typical errors." Explanations as to how English should not be used (We don't use *suggest* + indirect object + *to infinitive* . . .) are given there as bullet points consistently followed by examples of acceptable and unacceptable usage (e.g. He **suggested that** I should apply for a job . . . Not: ~~He suggested me to apply~~ . . .). Some grammar notes also feature shaded warning boxes with a series of examples of errors (Not: ~~We are used to go~~ . . . or ~~We use to go~~ . . . or ~~We were used to go~~ . . .).

Clearly, the presentation of examples showing errors in MLDs is widely varied. Yet, examples of errors are always visually enhanced. They are typically part of boxed warnings. Much less commonly do they follow a relevant sense.

Unfortunately, there is hardly any empirical research into the usefulness of examples of errors in MLDs. There is a study, though, which suggests that warning notes are not very effective in drawing learners' attention to errors. Chan (2012) asked 31 participants to use entries from CALD3 to judge sentence grammaticality. In addition to definitions and examples, some entries included "Common mistake" boxes with examples of errors. The research shows that the boxes were mostly ignored; they helped to make only one fifth of the participants' decisions. Far more were taken after reading examples (82%) or definitions (46%).

Importantly, the study was paper-based, and only four out of 10 CALD3 entries featured common mistake boxes. The role of examples showing errors as such was not investigated; the boxes just happened to be present in some entries. Besides, to gain an insight into the consultation process, self-reporting protocols (introspective questionnaires and think-aloud recordings) were used, followed by a post-task focus-group interview. Self-reporting has a potential for revealing fine details of dictionary use, but is open to several methodological problems. Instead of facts, beliefs and perceptions concerning the consultation process may be shared; subjects may wish to please the investigator or select what to report. Self-reporting performed during the task (like in the study in question) may also disturb look-up.

Warning notes including examples of errors have recently been studied by Nural, Nesi and Çakar (2022). Based on the analysis of LDOCE online, the authors suggested a typographical classification of warning notes into five categories and concluded that the use of typographical enhancements in each of them is very unsystematic. The warning categories also proved to be inconsistently employed, with hardly any direct connection with error types. The findings from the metalexicographic analysis were fed into an empirical study, in which four warning types were tested. The aim was to see if LDOCE warning notes help English learners correct errors and if their visual presentation affects success. The empirical study was naturalistic; 332 participants completed

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

an online survey with sentences to correct at their leisure, in the places of their choice and at their own pace. Most importantly, they could decide whether to click the supplied hyperlinks to LDOCE entries to do the task. Error correction rates ranged from 19% to 72%, with some dictionary conditions (dependent on warning types) more helpful than others. Unfortunately, real LDOCE entries prevented any tight control over important confounding variables like entry length or the position and length of warning notes. In the absence of log files, it was impossible to ascertain whether the entries were consulted and the warning messages (including examples of errors) noticed and read. In fact, information useful for error correction could also be found in other entry parts. Overall, the study suggests that typographical enhancements in warning notes (like borders or headings) and prominent positions might attract users' attention, but it does not show that warning notes affect users' performance. It only recognizes the potential of such notes for increasing the effectiveness of dictionary use, provided they are adequately conspicuous. The authors see a need for an experiment where purpose-built entries would be employed, with the length and placement of warning information strictly controlled.

Broadly speaking, the existing research suggests that warning notes are usually ignored (Chan, 2012) and need to be adequately highlighted to unlock their potential (Nural *et al.*, 2022). Yet, as real dictionary entries were drawn on, examples of errors were not isolated as a variable.

Findings pointing to the value of an inductive (example-driven) approach to learning lexico-grammatical patterns in CALL (e.g. Huang, 2014; Sun & Wang, 2003; Tsai, 2019) imply that examples of errors might benefit online dictionary users as a stand-alone dictionary feature rather than part of (usually neglected) warning notes. The latter often directly spell out prescribed language rules and patterns, and therefore embody the traditional deductive approach (rules first, then examples). Providing examples of errors without any explanatory notes could strengthen the inductive component in MLDs and encourage formulating underlying grammatical rules based on usage analysis. Such a maneuver might encourage usage-based learning and appears to be supported by important theories of L2 acquisition. For example, the depth of processing hypothesis predicts that more effective learning results from cognitively demanding activities that require manipulation of information input (Baddeley, 1997; Craik & Lockhart, 1972). Also, the constructivist learning theory justifies the expectation that more knowledge will be retained from individual data observation and analysis than explicit rule-driven instruction (Collentine, 2000; Resnick, 1987; cf. Cobb, 1999). While dictionary consultation may be considered to represent the deductive approach (Tsai, 2019), an inductive component based on learner-centered example analysis is inherent in dictionary use as well. After all, even online dictionaries are unable to explain *all* lexico-grammatical rules and patterns; many of them are just illustrated in corpus examples, leaving it up to the learner to infer rules. Examples of errors might help in this process. They may also provide a negative answer to some advanced learners' questions as to whether certain patterns are possible though not frequent enough to be included in a dictionary (cf. Section 1). Overall, it would be interesting to see if examples of errors unaccompanied by explicit warning notes help to rectify grammatical errors and learning usage.

3. Aim

The aim of the current study is to determine whether examples of errors in online dictionaries affect error correction accuracy as well as the immediate and delayed retention of usage. An attempt is also made to see if the position of examples showing errors in entries influences error correction and learning usage. Three research questions are posed:

Q1. Do examples of errors help to correct grammatical errors?
Q2. Is the immediate and delayed retention of usage affected by examples of errors?
Q3. Are error correction accuracy and usage acquisition conditioned by example positioning in entries?

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

4. Methods

4.1 Materials

A four-part online experiment was conducted to meet the aims of the study. It consisted of a pre-test, a main test, and immediate and delayed post-tests built around 18 sentences showing incorrect use of English. The English sentences with errors were drawn from Szymański (2006), a collection of sentence correction tests for upper-intermediate and advanced learners of English. The book contains sentences with the most typical grammatical errors made by Polish learners of English. Recourse to authentic learner language made it possible to design a more realistic online test, as advocated by Gyllstad and Snoder (2021). Only sometimes were the sentences slightly modified for length, proper names and infrequent words, which could affect comprehension. The incorrect structures from the sentences employed in the experiment are collated in Table 2, along with their corrections (see supplementary material). Several years of using the book by the author in teaching English to upper-intermediate and advanced students showed the errors to be the most difficult to rectify. As can be seen from the table, they concern various aspects of English grammar, including different errors in the test prevented the participants from predicting errors in consecutive sentences, becoming bored with the task as the study progressed and, eventually, losing concentration.

In the experiment, sentences with errors were displayed on separate webpages, and the errors were not highlighted. The aim of the pre-test was to find out if participants could identify and correct them without dictionary support. The post-tests checked the ability to correct the errors from memory immediately after the main test and two weeks later. Neither of them was announced in advance so as not to reinforce learning. They attempted to capture learning grammatical form, not meaning. In other words, the focus was on acquiring the grammatical form of the fairly entrenched constructions listed in Table 2, not the semantics of specific forms (e.g. the specific sense of the correct preposition). At each stage of the experiment, the sequence of the sentences was randomized. Having typed and submitted a correction, a participant could not change it.

In the main test, error correction was supported by online dictionary consultation. Each sentence with an error was followed by a monolingual dictionary entry compiled for the purpose of the study based on MLDs online. Two test versions were created, which differed only in the supplied dictionary entries. In one version, the entries offered regular examples. In the other, they additionally showed examples of errors. The entries were otherwise identical (see Appendices A and B in the supplementary material). The number of examples and their positioning were strictly controlled. In an entry, there were always six regular examples, only one of which was useful for error correction. In each test version, the regular examples helpful in the task at hand occupied entry-initial (1–2), medial (3–4) and final positions (5–6); that is, they were distributed in the first, second and third terciles in equal measure. In the test version with negative evidence, they were immediately followed by corresponding examples of errors. There was only one example showing an error in an entry. The information needed to correct the sentences could be extracted only from the relevant regular examples and examples of errors, and not from any other entry components.

Regular examples were drawn from MLDs. Examples of errors were created on the basis of the regular ones relevant to the error correction task and always showed exactly the errors that needed to be corrected in the test. They were printed in red and preceded by the note "✗*Don't say:*", as is often the case in LDOCE.[5] To illustrate, the regular example in the entry for *apologize*, *I think you should apologize to your brother*, taken from LDOCE online, was followed by the corresponding

[5]The red color was selected also because it is more salient than other colors (e.g. Gelasca *et al.*, 2005) and attracts attention most strongly in applied contexts (Fraser & Banks, 2004).

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

example showing an error: ✗*Don't say: I think you should apologize your brother*.[6] The example includes the error present in the sentence to be corrected: *The shop manager didn't even apologize us for the rude behaviour of his staff*. The strategy of supplying pairs of parallel regular examples and examples of errors was modeled on LDOCE, where the entry for *apologize* gives the following example: *We apologize to passengers for the delay*, and the corresponding example of an error: ✗*Don't say: We apologize passengers for the delay*.

4.2 Subjects and procedures

A total of 196 learners of English (B2 CEFR) at Adam Mickiewicz University participated in the study. Their proficiency level was established on the basis of the practical English exam taken at the end of the academic year as well as the course materials used in class throughout the whole year of instruction.[7] Thus, in Thomas's (2006) categorization, the institutional status of the participants (membership in a course group) and in-house assessment (proficiency defined by locally developed instruments) determined proficiency. There were 102 students who consulted entries with examples of errors, and 94 consulted those with regular examples only.

First, the participants were given the pre-test, which showed the incorrect sentences to be corrected without reference to any sources. Any cases where the errors were successfully corrected were excluded from further analysis[8]; 20 minutes were allotted to this stage. The pre-test was immediately followed by the main test, where the sentences had to be corrected based on dictionary consultation. One subject was assigned to one experimental condition – that is, obtained the test version either with or without examples of errors. In other words, some students in a class were randomly assigned to the test version where entries showed both regular examples and examples of errors, and the other students in the class were given the test version which offered only regular examples. The participants had 40 minutes for the task. They could ask for more time, but none of them voiced such a need. The post-test, which took place immediately after the main test, lasted 20 minutes, and so did the one administered two weeks later. In each of them, the sentences had to be corrected from memory.

4.3 Scoring

Sentence corrections that involved the target constructions and were linguistically appropriate were given one point. Any spelling errors insignificant to the meaning of the sentence (e.g. *appologise*, *familliar*, *synonymus*) were not taken into account. Attempts at correction in which the target structures were used inappropriately (e.g. *apologize for/at us*, *divorced to/against his wife*, *familiar by me*) or which involved other (non-target) parts of the sentences scored no points. The infrequent cases (0.9%) where instead of the target words the participants used alternative words/phrases or otherwise rephrased the sentences were counted as unsuccessful, too. For example, *separated from* used in place of *divorced with*, even though grammatically correct, was not accepted as it was semantically inappropriate and did not involve the target verb *divorce* in the right pattern. It did not follow from the consultation of the supplied dictionary entry for *divorce*, either.

The participants' answers were evaluated independently by two raters: an L1 English speaker and a speaker of English as an additional language (C2 CEFR) teaching EFL at university level. The

[6]This entry layout also justifies focusing on form in the study; "when learners expend the limited attentional resources processing the visually enhanced grammar features, they may . . . not have sufficient cognitive resources available for meaning construction" (Park *et al.*, 2012: 249).

[7]The experiment was conducted towards the end of the academic year, just before the final exam, when all the English course material had been covered. None of the participants failed the B2 test at the end of the course.

[8]There were only four correct responses in the pre-test; two participants corrected one error each, and one participant was able to correct two errors.

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

inter-rater agreement was 98.6%. After each rater had independently evaluated the responses, any cases where one accepted a correction and the other did not were thoroughly discussed in a face-to-face meeting, and a consensus was reached. For example, one rater rejected the response "*What did she look like*?" as a correction of "*How does she look like*?" on account of a different tense. The other one accepted it nonetheless. In the ensuing discussion, the raters agreed to accept the correction, as the target structure was there, and the use of tenses was not tested in the experiment. In a different case, the response "*I would rather not be arguing the point at the moment*" was turned down by one rater, since the participant not only corrected the negation but also used the continuous infinitive instead of the bare infinitive present in the original sentence: "*I wouldn't rather argue the point at the moment*." The other rater gave this correction one point because the target error (negation) was properly corrected, and the continuous infinitive did not affect the correction. Possibly, the adverbial of time (*at the moment*) simply made the subject opt for the continuous infinitive. At the meeting, the first rater conceded that the negation had been appropriately rectified and agreed to grant the response one point too.

5. Results

There were two factors in the study: *example type* (examples of errors vs. regular examples) and target *example position* in the entry (initial, medial and final). The former was a between-groups factor (categorical predictor) because each subject had access to only one dictionary version: either with or without examples of errors. The latter was a within-subject factor; each participant consulted entries where relevant examples were given in three places in equal measure. 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 2 x 3 between-within multivariate analysis of variance (MANOVA) was conducted.

5.1 Error correction accuracy in the main test, immediate and delayed retention

For each dependent variable at each level of the categorical predictor (examples of errors and regular examples), normality was determined based on the Kolmogorov–Smirnov and Lilliefors test for normality and the Shapiro–Wilk *W* test. For none of them was the assumption of normality violated ($p > .05$).

MANOVA results reveal a statistically significant main effect of example type (Wilks's lambda = 0.35, $F = 4.93$, $p = .032$, partial $\eta^2 = 0.649$). Neither the main effect of example position nor the interaction between the position of examples and their type was statistically significant ($p > .05$; see Table 3 in the supplementary material).

To explore the significant main effect (*example type*), univariate ANOVAs were conducted for each dependent variable. The Bonferroni correction was introduced to adjust *p* values because of the increased risk of Type I error when making multiple statistical tests.[9] The results reveal a significant difference in error correction accuracy in the delayed retention test between the groups consulting examples of errors and regular examples ($F_{(1,10)} = 11.6$, Bonferroni corrected $p = .020$; see Table 4 in the supplementary material). Examples of errors had a significant and positive effect on delayed retention. The subjects exposed to them two weeks earlier corrected from memory over 160% more errors (32.87%) than those who had not seen them in the main test (12.50%; 32.87*100/12.50 = 262.96; see Figure 1). The effect was quite large; over half of the variance in delayed retention can be attributed to example type (partial $\eta^2 = 0.537$).

[9]Obtained *p* values were multiplied by the number of conducted comparisons and compared against the alpha level of .05 (cf. Bland & Altman, 1995).

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

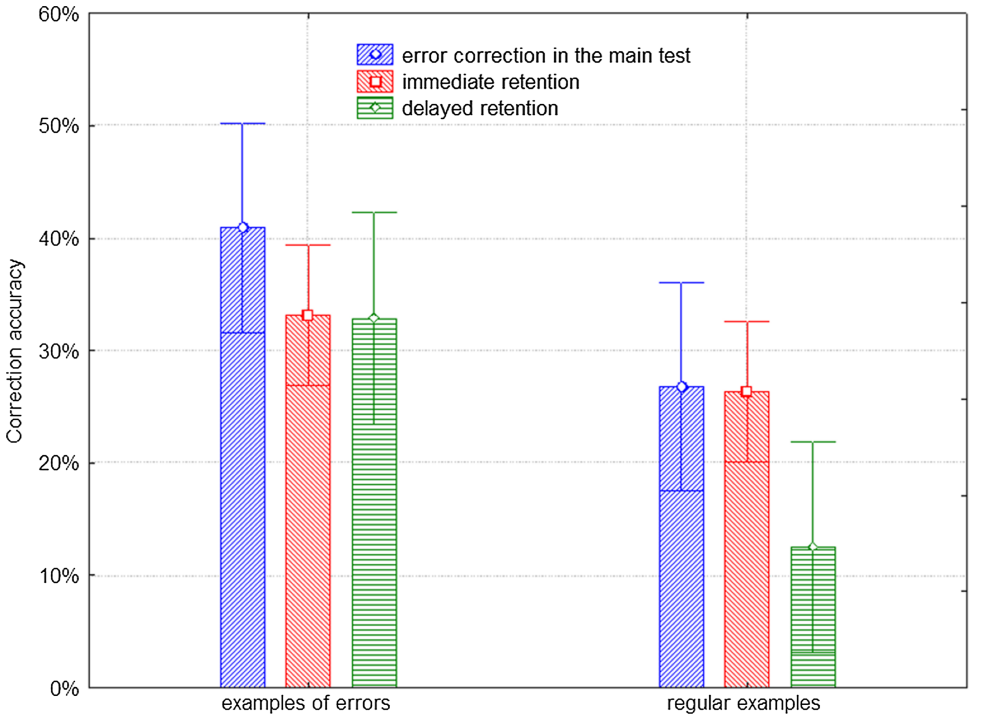


Figure 1. Error correction accuracy in the main test and immediate and delayed post-tests by example type. Vertical bars denote 95% confidence intervals

In the main test, examples of errors (40.90%) helped to correct over 50% more errors than regular ones (26.77%; 40.90*100/26.77 = 152.78). Yet, the effect was not statistically significant ($F_{(1,10)} = 5.8$, Bonferroni corrected $p = .112$, partial $\eta^2 = 0.366$). Finally, examples of errors did not significantly affect the ability to correct errors immediately after exposure (examples of errors: 33.17%; regular examples: 26.36%). Immediate retention was about one fourth better after reference to examples of errors (33.17*100/26.36 = 125.83), but the effect was not statistically significant ($F_{(1,10)} = 2.9$, Bonferroni corrected $p = .354$, partial $\eta^2 = 0.226$).

Interestingly, Figure 1 suggests that delayed retention was enhanced by examples of errors (32.87%) so much so that it virtually remained at the same level as immediate retention (33.17%). No such effect can be seen for regular examples, where delayed retention (12.50%) was much lower than immediate retention of usage (26.36%). To further investigate the contribution of examples of errors to delayed retention, multivariate tests of significance were performed. An attempt was made to see if the differences between immediate and delayed retention in the groups consulting examples of errors and regular examples were statistically significant.

The results show that the difference of around 1% between the immediate and delayed retention results among the participants consulting examples of errors was not statistically significant (Wilks's lambda = 1.00, $F = 0.00$, $p = .952$; 32.87*100/33.17 = 99.10). The corresponding difference of about 50% among the participants using only regular examples was statistically significant (Wilks's lambda = 0.55, $F = 8.19$, $p = .017$; 12.50*100/26.36 = 47.42). It follows that in the absence of examples showing errors in entries, dictionary-based retention of correct usage deteriorates significantly over time (from 26.36% to 12.50% in the current study). However, when examples of errors are introduced into entries, delayed retention of usage remains at the same level as immediate retention (around 33%).

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

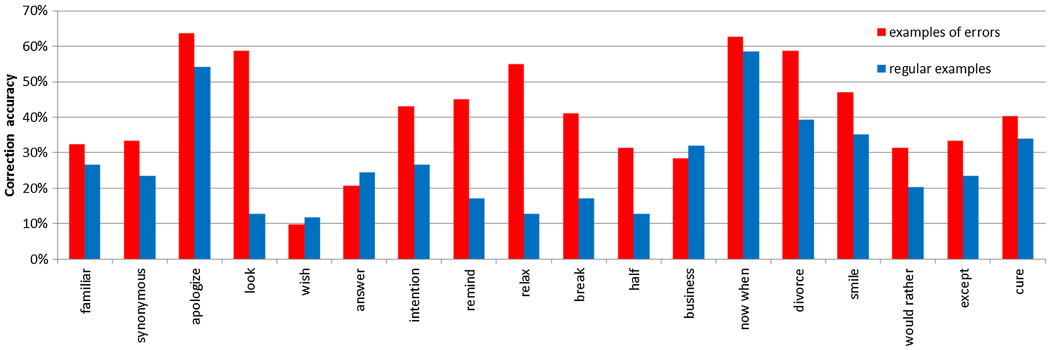


Figure 2. Error correction accuracy in the main test by error and example type

5.2 Ancillary findings

In the current investigation, different errors were tested, and their categories were not controlled. There were errors in prepositions (**familiar for sb*), unreal past (**I wish I would*), verb patterns (**remind sb of doing sth*), adjectives (**ten-minutes break*), verb+noun collocations (**make business*) or conjunctions (**now when*). To see whether some errors were more difficult to handle than the others, correction accuracy in the main test was examined for each of them (Figure 2).

As can be seen, for both groups, **apologize sb* and **now when* were the easiest to correct, and **I wish that I would* was the most challenging. Except for three cases (**I wish that I would*, **an answer for a question*, **make business*), error correction proved simpler in the presence of examples showing errors. Their greatest advantage was observed for **How does sth look like?* and **relax yourself*, where they increased correction accuracy by about three and a half times (*look*: 58.82*100/12.77 = 460.78, *relax*: 54.90*100/12.77 = 430.07). The contribution of examples of errors was the smallest for **now when*, where it approximated only 7% (62.75*100/58.51 = 107.24), followed by **apologize sb*, where it reached 17% (63.73*100/54.26 = 117.45). It is interesting to note that in all these cases, the correct Polish constructions mirror the erroneous English usage: **how does it look like – jak to wygląda*, **relax oneself – relaksować się*, **now when – teraz kiedy*, **apologize sb – przepraszać kogoś*. Thus, examples showing errors proved both most and least useful when interference from L1 could hinder error correction. It follows that grammatical incongruence between the languages is unlikely to have affected the usefulness of such examples. In the three cases where better error correction results were obtained in the absence of examples of errors, the differences in favor of regular examples ranged from 12% to about 20% (*business*: 31.91*100/28.43 = 112.25, *answer*: 24.47*100/20.59 = 118.84, *wish*: 11.70*100/9.80 = 119.36).

5.3 Summary of the results

The results show that examples of errors in online dictionaries do not significantly increase error correction accuracy (Q1). They do not affect the ability to remember usage immediately after exposure either, but contribute to its retention in the long run (Q2). Examples of errors actually prevent forgetting usage over time and keep delayed retention at the same level as immediate retention. Finally, the place of examples in entries does not influence either error correction success or the retention of grammatical forms (Q3).

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

6. Discussion

The study does not reveal a statistically significant effect of examples of errors on correction accuracy during dictionary consultation. Yet, it does show that such examples helped to correct over 50% more errors in the main test than regular ones, which has undeniable *practical* significance. The difference may be attributed to the typography and the layout of the entry. It is likely that the red color of examples of errors alerted the participants to the constructions indicated as wrong and helped identify errors in sentence cues. Moreover, examples of errors were always preceded by regular ones showing correct usage (Section 4.1). Such a predictable layout may have helped work out how to correct errors.

The effort invested in analyzing examples of errors, comparing them with corresponding regular examples and sentence cues, apparently strengthened the memory trace, which might explain better delayed retention. Possibly, the participants remembered examples of errors from the main test, where error correction was supported by dictionary entries, and used this information to guide their later performance. Having already been warned against specific errors, they found it easier to recognize them in sentence cues two weeks later and come up with corrections. The significant boost in delayed retention of as much as 160% might also result from the potential of stand-alone examples of errors for streamlining inductive reasoning. Showing what is incorrect and should be avoided, they improve the accuracy of induction, but naturally require additional mental exertion and increase the depth of cognitive processing, which might result in better retention of knowledge in the long run (cf. Baddeley, 1997; Collentine, 2000; Craik & Lockhart, 1972; Resnick, 1987). Besides, in entries with examples of errors, target structures were addressed twice: in regular examples and in examples of errors, which might produce synergy effects. Such a juxtaposition might contribute to conscious noticing, which is conducive to learning (Schmidt, 1990). Finally, the signaling function of red is reflected in human memory. Typical signal colors, like red, stick in later memory better than non-signal ones and help to retain the significance of objects (Kuhbandner, Spitzer, Lichtenfeld & Pekrun, 2015). Thus, the chances are that red examples of errors were more successfully stored in later memory, which further contributed to better delayed retention. Revealing how visually enhanced examples of errors add to remembering grammatical forms in the long term is an important contribution of the present study to the field, as hardly any investigation into VIE undertakes to determine the durability of VIE benefits (Lee & Huang, 2008; Park *et al.*, 2012).

Finally, the observation that the place of examples in entries has no significance for error correction success and the retention of grammatical forms diverges from the findings that point to the salience of entry top or bottom (cf. Section 1). Yet, previous conclusions concern the positioning of senses and collocations and its role in different tasks, and may not be generalizable to multiple examples in entries. Maybe within an example set there are no more or less salient positions, but what matters is the content of examples or their conspicuity.

7. Limitations

To eliminate distraction caused by different contexts of errors, the same sentence cues were used throughout. Although the sequence of the sentences was changed at each stage of the experiment (Section 4.1), the subjects may have remembered the specific errors in the particular contexts. It is not known if they learned to generalize.

The entries compiled for the study were monosemous. However, entry length could have an effect on the noticeability of examples of errors; they might be more difficult to retrieve from polysemous entries.

As error types were not a factor in the experiment, it is impossible to conclude that some errors need examples more than others. Yet, the item analysis (Section 5.2) indicates that the contribution of examples of errors to correction accuracy is by no means uniform.

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

No tools were employed to measure noticing examples of errors. The research aimed to determine whether entries with examples of errors affect correction accuracy and the retention of usage. It was not designed to gain an insight into noticing such examples and the amount of attention paid to them. Thus, while the experiment makes it possible to evaluate the contribution of different entry formats (with and without examples of errors) to correction accuracy and the retention of grammatical forms, it does not disclose details of the dictionary consultation process and the actual role of the perceptual prominence of examples of errors in drawing attention. Yet, attending to examples of errors surely warrants careful consideration in future studies.

8. Further research and lexicographic recommendations

Noticing examples of errors in dictionaries might be an interesting area of further investigation. Considering the limitations of think-aloud protocols (cf. Section 2) and the fact that they tend to investigate the awareness of VIE rather than real attention devoted to it and maintained for some time (Loewen & Inceoglu, 2016), eye-tracking methodology, which measures attentional focus in real time, would be most recommendable to establish what exactly users attend to in an entry with examples of errors.

Eye tracking is also necessary to investigate salient positions of examples showing errors. It is not known if consecutive examples are scanned top-down or bottom-up, or maybe the visual salience of examples of errors gives them priority and makes users hone in on them. Polysemous entries with more numerous example clusters should also be examined. Maybe in longer entries with more examples users tend to look primarily at those given at the top or bottom.

Simard (2009) argues that the selection of typographic cues should not be random. It might be interesting to see how different highlighting methods (e.g. strikethrough font, capital letters or color) or their combinations draw users' attention to examples of errors. It is by no means certain either that red serves better than other colors to bring out such examples. It conventionally performs a signaling or warning function, indicates importance, danger or caution. However, when used to mark errors in essays, it makes learners feel abused and evokes anger and sadness (Dukes & Albanesi, 2013). Such voices against using red to highlight errors might be considered in lexicography.

It might be worthwhile to investigate error types whose correction is facilitated most by examples of errors. The item analysis (Section 5.2) shows that although examples of errors were helpful in the majority of cases, their contribution varied. There might be error categories for which such examples are most recommendable.

It might be instructive to compare the usefulness of stand-alone examples of errors to those in warning boxes. It is not known whether more deductive warning boxes, which give rules of grammar (Section 1), are more beneficial than stand-alone examples of errors, which encourage induction. It would also be recommendable to engage participants in less controlled computer-based tasks (e.g. free writing or proofreading) to find out if examples of errors help avoid errors in more naturalistic linguistic contexts.

To better cater for learners' L1-specific difficulties (cf. Section 1), examples showing errors in MLDs could be localized and bilingualized. Localizing such examples would require selecting them from a mono-L1 English learner corpus to highlight errors typical of that language community. Subsequent translation into L1 (or bilingualization) would assist less advanced learners and expose possible inter-language anisomorphism. As space constraints are much less of a problem in online dictionaries, examples of errors could be more numerous. Frankenberg-Garcia (2014: 140–141) holds that "it is only when we have access to a number of examples of the same kind that we are able to detect conventional patterns of use." Besides, examples of errors may go beyond issues in grammar or lexis. Spoken learner corpora reveal typical pronunciation errors. Voice recordings of such errors (made by learners from a given linguistic background) could be

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

juxtaposed with recordings of correct native-speaker pronunciation. Phonetic transcription could be added for more advanced learners. Thus, negative evidence in dictionaries could go multimodal (aural and written).

9. Conclusion

Overall, the current study exhibits statistically significant and positive evidence for including examples of errors in online learners' dictionaries. Their high pedagogical value is evident in the substantial contribution they make to retaining usage in the long run (cf. Tomasello & Herron, 1988, 1989). Since examples of errors successfully prevent dictionary users from forgetting fairly entrenched constructions over time, they appear to constitute a perfect complement to the dictionary microstructure. Keeping the long-term retention of usage at the level of immediate retention, examples showing errors turn dictionaries into valuable learning tools, and possibly give them a competitive edge over Google or ChatGPT, increasingly often consulted to solve linguistic problems. The current investigation also indicates that about 50% more errors are corrected in the presence of examples of errors, which no doubt has *practical* (even if not statistical) significance. Examples of errors can thus be of practical use in various tasks that require error correction (or avoidance), like completing written assignments or proofreading. Importantly, they are demonstrated to be a valuable stand-alone, induction-oriented entry feature, not enclosed in (oft-ignored) descriptive warning boxes. Overall, as long-term retention boosters and practical support in error correction, examples of errors should find their permanent place in MLDs.

Supplementary material. To view supplementary material referred to in this article, please visit https://doi.org/10.1017/S0958344023000241

Ethical statement and competing interests. The data gathered in the study were completely anonymized. All the participants were adult university students and gave prior informed consent to participate. Participation was voluntary, and participants were given the opportunity to withdraw at any time. The experiment constituted an extension of in-class activities, so inflicting any psychological damage on the participants is highly unlikely. The participants were treated in accordance with the provisions of Polish law. The author declares no competing interests.

References

Baddeley, A. D. (1997) *Human memory: Theory and practice* (Rev. ed.). Hove: Psychology Press.

Bland, J. M. & Altman, D. G. (1995) Multiple significance tests: The Bonferroni method. *The British Medical Journal*, 310: 170. https://doi.org/10.1136/bmj.310.6973.170

Chan, A. Y. W. (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

Chen, Y. (2020) The efficacy of dictionary use for collocation learning: A case study of smartphone dictionary. *Lexicographical Studies*, 2: 20–31.

Cobb, T. (1999) Applying constructivism: A test for the learner-as-scientist. *Educational Technology Research & Development*, 47(3): 15–31. https://doi.org/10.1007/BF02299631

Collentine, J. (2000) Insights into the construction of grammatical knowledge provided by user-behavior tracking technologies. *Language Learning & Technology*, 3(2): 46–60.

Craik, F. I. M. & Lockhart, R. S. (1972) Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6): 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

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

Dukes, R. L. & Albanesi, H. (2013) Seeing red: Quality of an essay, color of the grading pen, and student reactions to the grading process. *The Social Science Journal*, 50(1): 96–100. https://doi.org/10.1016/j.soscij.2012.07.005

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. https://doi.org/10.1093/ijl/ecu012

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

Dziemianko, A. (2015) Colours in online dictionaries: A case of functional labels. *International Journal of Lexicography*, 28(1): 27–61. https://doi.org/10.1093/ijl/ecu028

Dziemianko, A. (2017) Dictionary entries and bathtubs: Does it make sense? *International Journal of Lexicography*, 30(3): 263–284. https://doi.org/10.1093/ijl/ecw010

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

Fraser, T. & Banks, A. (2004) *Designer's color manual: The complete guide to color theory and application*. San Francisco: Chronicle Books.

Gelasca, E. D., Tomasic, D. & Ebrahimi, T. (2005) Which colors best catch your eyes: A subjective study of color saliency. In *Proceedings of First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*. Washington: SPIE.

Granger, S. (2008) Learner corpora. In Lüdeling, A. & Kytö, M. (eds.), *Corpus linguistics: An international handbook: Volume 1*. Berlin: Walter de Gruyter, 259–275.

Gyllstad, H. & Snoder, P. (2021) Exploring learner corpus data for language testing and assessment purposes: The case of verb + noun collocations. In Granger, S. (ed.), *Perspectives on the L2 phrasicon: The view from learner corpora*. Bristol: Multilingual Matters, 49–71. https://doi.org/10.21832/9781788924863-004

Huang, Z. (2014) The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL*, 26(2): 163–183. https://doi.org/10.1017/S0958344014000020

Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. (1995) Does textual enhancement promote noticing? A think-aloud protocol analysis. In Schmidt, R. (ed.), *Attention and awareness in foreign language learning and teaching*. Honolulu: University of Hawai'i Press, 183–216.

Kuhbandner, C., Spitzer, B., Lichtenfeld, S., & Pekrun, R. (2015) Differential binding of colors to objects in memory: Red and yellow stick better than blue and green. *Frontiers in Psychology*, 6: Article 231. https://doi.org/10.3389/fpsyg.2015.00231

Lee, M. & Jung, J. (2021) Effects of textual enhancement and task manipulation on L2 learners' attentional processes and grammatical knowledge development: A mixed methods study. *Language Teaching Research*. Advance online publication. https://doi.org/10.1177/13621688211034640

Lee, S.-K. (2007) Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning*, 57(1): 87–118. https://doi.org/10.1111/j.1467-9922.2007.00400.x

Lee, S.-K. & Huang, H. (2008) Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition*, 30(3): 307–331. https://doi.org/10.1017/S0272263108080479

Lew, R., Grzelak, M. & Leszkowicz, M. (2013) How dictionary users choose senses in bilingual dictionary entries: An eye-tracking study. *Lexikos*, 23: 228–254. https://doi.org/10.5788/23-1-1213

Li, L. & Xu, H. (2015) Using an online dictionary for identifying the meanings of verb phrases by Chinese EFL learners. *Lexikos*, 25: 191–209. https://doi.org/10.5788/25-1-1295

Loewen, S. & Inceoglu, S. (2016) The effectiveness of visual input enhancement on the noticing and L2 development of the Spanish past tense. *Studies in Second Language Learning and Teaching*, 6(1): 89–110. https://doi.org/10.14746/ssllt.2016.6.1.5

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.

Park, H., Choi, S. & Lee, M. (2012) Visual input enhancement, attention, grammar learning, and reading comprehension: An eye movement study. *English Teaching*, 67(4): 241–265. https://doi.org/10.15858/engtea.67.4.201212.241

Resnick, L. B. (1987) Learning in school and out. *Educational Researcher*, 16(9): 13–20, 54. https://doi.org/10.2307/1175725

Rundell, M. (1999) Dictionary use in production. *International Journal of Lexicography*, 12(1): 35–53. https://doi.org/10.1093/ijl/12.1.35

Schmidt, R. W. (1990) The role of consciousness in second language learning. *Applied Linguistics*, 11(2): 129–158. https://doi.org/10.1093/applin/11.2.129

Simard, D. (2009) Differential effects of textual enhancement formats on intake. *System*, 37(1): 124–135. https://doi.org/10.1016/j.system.2008.06.005

Sun, Y.-C. & Wang, L.-Y. (2003) Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, 16(1): 83–94. https://doi.org/10.1076/call.16.1.83.15528

Szymański, J. (2006) *Sentence correction tests*. Włocławek: Altravox Press.

Thomas, M. (2006) Research synthesis and historiography: The case of assessment of second language proficiency. In Norris, J. M. & Ortega, L. (eds.), *Synthesizing research on language learning and teaching*. Amsterdam: Benjamins, 279–298. https://doi.org/10.1075/lllt.13.13tho

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press

Tomasello, M. & Herron, C. (1988) Down the garden path: Inducing and correcting overgeneralization errors in the foreign language classroom. *Applied Psycholinguistics*, 9(3): 237–246. https://doi.org/10.1017/S0142716400007827
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
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
Winke, P. M. (2013) The effects of input enhancement on grammar learning and comprehension: A modified replication of Lee (2007) with eye-movement data. *Studies in Second Language Acquisition*, 35(2): 323–352. https://doi.org/10.1017/S0272263112000903
Wong, W. (2003) Textual enhancement and simplified input: Effects on L2 comprehension and acquisition of non-meaningful grammatical form. *Applied Language Learning*, 13(2): 17–45.

Dictionaries

[CALD] *Cambridge advanced learner's dictionary*. https://dictionary.cambridge.org/pl/dictionary/english
[LDOCE] *Longman dictionary of contemporary English*. https://www.ldoceonline.com/
[MDO] *Macmillan dictionary online*. https://www.macmillandictionary.com/
[OALD] *Oxford advanced learner's dictionary*. https://www.oxfordlearnersdictionaries.com

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.

Author ORCiD. Anna Dziemianko, https://orcid.org/0000-0002-6325-2948

https://doi.org/10.1017/S0958344023000241 Published online by Cambridge University Press