OXFORD

Article

WARNING NOTES IN A LEARNER'S DICTIONARY: A STUDY OF THE EFFECTIVENESS OF DIFFERENT FORMATS

Sukru Nural (iD)

MEF University, Turkey (nurals@mef.edu.tr)

Hilary Nesi

Coventry University, United Kingdom (h.nesi@coventry.ac.uk)

Tuna Cakar

MEF University, Turkey (cakart@mef.edu.tr)

Abstract

This study used an online correction task to explore the extent to which different types of warning notes in *Longman Dictionary of Contemporary English Online* were heeded when users tried to correct errors in the use of L2 target words. The task was completed by 332 participants, yielding 1,819 answers produced after clicking on links to relevant entries. Warning notes were categorised in terms of their formatting features, but there were found to be inconsistencies in the way the dictionary associated different categories with different kinds of learner error. Participants judged warning notes with more visual enhancements to be more useful, but in the correction task the position of the warning notes also seemed to affect the degree to which the warnings were successfully applied. Different types of warning notes in learners' dictionaries have not been examined previously in any depth, and the results suggest that some adjustments to formatting and placement might make them more effective.

Key words: warning notes, typographical features, noticing, error correction, dictionary use

1. Introduction

In second language acquisition studies it is generally accepted that visual cues in the text can encourage noticing and that input must be noticed before it can be processed for acquisition (Schmidt, 2001). A number of focus-on-form intervention studies have examined the effect of various kinds of visual enhancement on grammar learning, as compared to explicit

instruction using metalinguistic terminology, and/or 'input flooding' (exposure to multiple examples of the target structure). Results have been inconclusive, but a meta-analysis conducted by Lee and Huang (2008) suggests the possibility of a small positive effect for visual enhancement. Lee and Huang point out that, as in all pedagogical interventions, this effect is likely to be greatest when learners are already somewhat familiar with the target forms.

However, although visual enhancement may increase the salience of learning material, some types of salient presentation (for example placement in a position where it is likely to be noticed) would seem to make information easier to access, while other types of presentation (such as non-standard fonts and typefaces) make information access slightly more effortful. The theory of 'deep processing' argues that greater learning results if content is harder to obtain, because this encourages engagement in more active learning strategies (Craik and Lockhart 1972, Craik and Tulving 1975). Hulstijn and Laufer's (2001) Involvement Load Hypothesis makes a similar argument with reference to vocabulary retention. The positive effects of deep processing seem to be supported by studies such as those of Diemand-Yauman et al. (2011), who found that university students retained factual information better if it was presented in slightly harder-to-read fonts (grey-scale Comic Sans MS and Bodoni MT), and Peters (2012) who found that underlined and bold typeface had a positive effect on language learners' recall of formulaic sequences. Diemand-Yauman et al. (2011: 111) concluded that 'superficial changes to learning materials could yield significant improvements in educational outcomes'.

The idea that presentation might affect the way we attend to information is probably particularly relevant to e-dictionary design and use. Educators have long suspected that the speed and ease of e-dictionary consultation impacts negatively on retention (Taylor and Chan 1994, Sharpe 1995, Zhang 2004, Stirling 2005), and recent psychological, psychiatric and neuroimaging research indicates that accessing information online has an adverse effect on our attentional capacities (Firth et al. 2019).

Warning notes in learners' dictionaries contain the kind of 'negative input enhancement' described by Sharwood-Smith (1993: 177). In these notes language errors of various kinds, usually identified through analysis of learner corpus data, are flagged by means of multiple non-standard typographical features which differ from the surrounding text, such as boldfacing, capitalization, italicization, strikethrough and colour-coding, in an attempt to increase their salience and encourage learners to take note. Warning notes are frequently (but not always) placed within text boxes, a format which might also provide a visual cue for users. Boxes may be placed at the end of the entire entry, or immediately following an entry component or a particular polysemous sense, depending on the scope of the message. Gouws and Prinsloo (2010) describe how text boxes in the *Macmillan English Dictionary* and in the Afrikaans-English section of the *Reader's Digest Dictionary* are often used to flag up contrasts between similar words, thus helping the user avoid incorrect use. They advise lexicographers to reserve text boxes for 'data in need of a position of salience' (2010: 509), but argue that their judicious use 'can really enhance the data transfer in dictionaries' (2010: 510).

As far as we are aware no prior empirical studies have focused explicitly on learners' use of dictionary warning notes, although there has been research into the effect on dictionary use of typographical enhancements. An eye-tracking experiment by Lew et al. (2013), for example, found that elements in bold in dictionary entries caught the attention of users, and Dziemianko (2014) found that study participants were far more likely to fill gapped

sentences with correct collocations if the relevant information was provided in bold rather than in boxes. Dziemianko (2015) investigated the effect of using colour in part-of-speech and syntactic labels, and found that participants retrieved the grammatical information they needed more quickly, and applied it more correctly, if the labels were in colour. Dai et al. (2019) found that salient positioning (right under the headword) resulted in faster access to conceptual metaphor information.

However, studies which have recorded what learners notice during dictionary consultation suggest that warning notes might not be very effective in drawing learners' attention to errors. Chan (2011), for example, asked participants to use entries from *Cambridge Advanced Learner's Dictionary* (Third Edition) to judge sentences grammatically. In addition to definitions and examples, some of the entries containing the target words included boxed warning messages which identified and corrected common mistakes, yet only 21.3% of the participants' decisions were made after consulting these messages, and far more were made after consulting examples (82%), or definitions (46%). Moreover, when Chan (2012) examined learners' use of four monolingual learners' dictionaries she found that participants very rarely reported that they had noticed this extra visually enhanced information, claiming to have referred instead to the information contained in examples and definitions.

This study will examine the effect of four different types of visually enhanced input in dictionary warning notes, in an attempt to determine the extent to which warning notes are actually noticed and applied by learners, and whether some types of enhancement are more useful than others. It is not concerned with the acquisition and retention of lexical knowledge but it is concerned with noticing, a pre-requisite both for vocabulary learning and for successful dictionary consultation.

The free online version of *Longman Dictionary of Contemporary English* was used for the study; the types of warning notes it contains will be described in the following section.

2. Background

2.1. Warning notes in *Longman Dictionary of Contemporary English Online*

Warning notes can be compared to road signs which ask drivers not to do something (e.g. 'Don't exceed the speed limit') as well as, or instead of, telling them what they should do (e.g. 'Reduce speed limit now'). They provide an example of acceptable usage alongside one of unacceptable usage, which in this paper we will call a 'non-example'. In their review of *LDOCE6*, Osada et al. (2015: 64) refer to the use of "✗ Don't say" warnings in grammar boxes, but in fact not all *LDOCE6* grammar boxes contain a warning not to do something, and not all warnings are placed inside boxes. Moreover not all *LDOCE6* warning notes are classed as dealing with grammatical points; some are marked as dealing with the way in which the item is used in combination with other words (collocations), and/or alert dictionary users to word pairs which are confusable, either orthographically or semantically (e.g. PLAY/PERFORM).

The warning notes in *LDOCE Online* can be classified into four major types according to the typographic enhancements they employ, as shown in Table 1. They differ in minor ways from *LDOCE6* in print in that warnings marked with a red triangle online are marked with a cautionary exclamation mark in the print version, and the print version frames more of its warning notes within boxes.

Table 1: Typographical features of warning notes in *LDOCE online*

	Typographical features	Type 1	Type 2A	Type 2B	Type 3	Type 4
1	**Heading** (lower case)		✓			
2	**HEADING** (upper case or lower case)			✓	✓	
3	non-example (in red)		✓	✓	✓	
4	grey		✓	✓	✓	
5	□ (enclosed in a Box)		✓	✓		✓
6	✗Don't say: (in red)		✓	✓		
7	✗Don't say:				✓	✓
8	**Sub-heading**			✓	✓	
9	▶ (in red)	✓				✓
10	*italics*	✓				✓
11	HEADING (capital letters against a purple background)					✓
12	**bold**	✓				
13	~~strikethrough~~					✓
14	*(NOT...)*	✓				

It seems to be the intention in LDOCE that different formats should be used for warnings concerning different language areas, Type 1 being used to describe verb patterns and phraseology (or information about similar spellings or pronunciations), Types 2A and Type 2B to highlight points of grammar, Type 3 to deal with usage (mostly in relation to the use of function words), and Type 4 for collocational information. However, one important initial finding was that there was a good deal of overlap between the points dealt with by these five types, and that it was not always obvious why one type of warning note had been chosen rather than another. Issues of verb patterning, grammar, usage and collocation frequently intertwine, and the lexico-grammatical nature of language errors means that in practice it is difficult to divide them neatly into categories. Although they are formatted somewhat differently, Types 2A and 2B both deal with the same sort of errors, and it is sometimes difficult to distinguish between these 'grammar' errors and Type 1 errors, or Type 3 errors involving function words. Type 1 warning notes also sometimes stray into the area of collocational choice, as in the entry for PERFORM: "Do not use **perform** to say what person an actor pretends to be in a play, film etc. Use **play**: *John Wayne played (NOT performed) a Roman soldier in the film*". In the entry for TEST (noun), both Type 1 and Type 4 notes are used in the same way to warn users about the choice of the accompanying verb.

For ease of reference we have numbered the four warning types we have identified according to the order in which they might appear in entries in *LDOCE Online*. If there is more than one warning note in an entry, Type 1 will always appear first, Types 2A and 2B will occur before Types 3 and 4, and Type 3 will always appear before Type 4. Types 2A

and 2B are not distinguishable in terms of position or content, but are treated as separate types in this study because of differences in their headings, as shown in Table 1.

Different word classes with the same form may have warnings of different types, as in the entry for PAY, with a Type 2B warning for the verb and a Type 4 warning for the noun. Sometimes the same lemma is given more than one warning; for example REALIZE is given a Type 1 warning regarding the *about/of* pattern, and (further down the entry) a Type 2B warning not to use the progressive. Similarly GET is given a Type 2A warning not to use the progressive followed by a Type 2B warning not to use the passive.

Table 1 shows that Type 2B and Type 4 have the greatest number of typographical enhancements (six), followed by Type 2A and Type 3 (five), and finally Type 1 (four). The reasons for these formatting differences are not obvious. The types of warning enclosed in boxes (Types 2A, 2B and 4) do not seem to deserve to be made more salient than those in unboxed warning types, for example. Similarly the use of variously formatted headings and subheadings does not seem to reflect the relative importance of the message. For example the warning not to use a particular word or structure is conveyed in Types 2A, 2B, and 3 by a '✗ Don't say' message in either black or red type followed by an erroneous non-example in red type, but the same message in Type 4 is conveyed by a non-example on its own, in black strikethrough. There does not seem to be any clear reason for these differences.

The common errors treated in warning notes in the Longman dictionaries come from the *Longman Learner's Corpus*, which, according to the Pearson website, contains 12 million words of students' essays and exam scripts. In order to choose the warning notes for our study we looked at the core vocabulary which is marked in *LDOCE6* and *LDOCE Online*, drawn from the *Longman Communication 9000* (*LC9000*). Words in this list are divided into three frequency bands: 3,049 words to represent the most frequent words in spoken and written English, 3,104 to represent mid-frequency words, and 2,874 to represent lower-frequency words. Although derived forms may be grouped within the same band in *LC9000* (e.g. REQUIRE-REQUIREMENT; SUGGEST-SUGGESTION), this is not always the case. For example DESIRE and VALUE are in the high frequency band as verbs but in the medium frequency band as nouns.

Of the 9,027 words in *LC9000*, 6.3% are given warnings in *LDOCE Online*. Table 2 categorises the 575 warning notes provided for these words. As indicated in Table 2, the higher the frequency band the greater the number of warning notes.

The following section describes in more detail the types of warning notes in *LDOCE Online*.

2.1.1. *The four different types of warning notes*
Type 1:
Type 1 occurs immediately after the relevant sense, and is indicated by a red triangle (▶) as in Figure 1. This is followed by an example of the more accurate and acceptable form (in bold) and often an italicized note, which may be preceded by 'NOT' in capital letters. Probably the shortest Type 1 warning note is for the word FARM: ▶ Say on a farm not 'in a farm'.

Type 2A:
Type 2A warning notes are presented within a framed grammar box with a blue bold heading in small letters (Grammar) and no subheading (see Figure 2). The explanation supported by example sentences in grey is followed by a red cross (✗) and a red 'Don't say' note with an erroneous sentence again in a red font. The pattern of the headwords is

Table 2: Distribution of warning types for words in the *Longman Communication 9000*

Warning types	High frequency	Medium Frequency	Low Frequency	Total
Type 1	71	22	5	98
Type 2A	68	18	3	89
Type 2B	245 including 19 with more than one subheading	22 including 1 with more than one subheading	9	276
Type 3	25	2	None	27
Type 4	79	6	None	85
TOTAL:	**488**	**70**	**17**	**575**

o·bey /əʊˈbeɪ, ə- $ oʊ-, ə-/ ●●● W3   **verb [intransitive, transitive]** 🔊 🔊

to do what someone in authority tells you to do, or what a law or rule says you must do OPP **disobey**

🔊 The little boy made no effort to obey.

🔊 'Sit!' he said, and the dog obeyed him instantly.

obey an order/command

🔊 Soldiers are expected to obey orders without questioning them.

obey the law/rules

🔊 Failure to obey the law can lead to a large fine.

▶ Don't say 'obey to someone/something'. Say **obey someone/something**: *He refused to obey their orders (NOT obey to their orders).*

Figure 1: An entry containing a Type 1 warning note: OBEY (v.):

Grammar
You **accuse** someone **of doing** something:
   They accused him of cheating.
✗ Don't say: accuse someone to do something

Figure 2: A Type 2A warning note: ACCUSE *(v.)*

presented in bold. The notes can be very short: '✗ Don't say: in fire' is probably the shortest warning note in the dictionary.

Type 2B:
Like Type 2A, Type 2B warning notes are also presented within a framed grammar box and have a a blue bold heading, although in the case of 2B this heading is usually in capital letters (GRAMMAR). However, Type 2B warning notes also have a blue subheading (e.g. Prepositions with…) in small letters (see Figure 3). A metalinguistic explanation is supported by illustrative sentences in grey, and is followed by a red cross (✗) and a red 'Don't say' note with a non-example, again in red font. The headwords within this explanation are presented in **bold**.

GRAMMAR: Prepositions with married

- You are **married to** someone:
  He is married to an American.
  Last year I got married to Tom.
  X Don't say: He is married with an American. | Last year I got married with Tom.
- You use **married with** when talking about how many children someone has:
  Richard is now happily married with two young children.

Figure 3: A Type 2B warning note: MARRIED (*adj.*)

The length of Type 2B warning notes ranges from 4 to 180 words, and the same box may contain more than one subheading (e.g. 'Using the progressive' and 'Using the passive'), with Type 2B warning notes under each subheading or only under one (e.g. in the case of APPEAR, which has a warning note under the subheading 'Using the progressive' but a note without a non-example under the subheading 'Linking verbs'). A total of 20 words from the *Longman Communication 9000* (*LC9000*) have more than one Type 2B warning note under the same subheading. (e.g. AGREE, BELIEVE, DEPEND, DISCUSS, DISLIKE).

Type 2B warning notes have subheadings relating to verb patterns, prepositions, word order, the progressive aspect, countable and uncountable nouns, and tenses (see Table 3). They are presented after all the senses of the headword have been listed, rather than immediately after one relevant sense, but within the box senses may be referred to by their number, to indicate that the warning applies to a particular sense (e.g. 'Senses 1-4 of *have* are not used in the progressive'). If the entry includes a collocations box, 2B warning notes will always precede this.

Type 3:
Type 3 warning notes are presented within a framed box, with a blue bold USAGE heading in capital letters. The note makes a metalinguistic comparison of two or more grammatical

Table 3: Subheadings for TYPE 2B across three bands in the *Longman Communication 9000*

TYPE 2B subheadings	High-frequency	Mid frequency	Low-frequency	Total
Patterns with…	88	5	2	95
Using the progressive	41	10	7	58
Comparison	41	2	None	43
Prepositions with…	17	1	None	18
Word order	18	1	None	19
Order of adjectives	11	None	None	11
Choosing the right tense	10	None	None	10
Negatives	10	None	None	10
Singular or plural verb?	7	None	None	7
Countable or uncountable?	6	None	None	6
Comparatives	5	None	None	5
Using the passive	3	None	None	3
Linking verbs	3	None	None	3
Spelling	1	None	None	1
Possessives	1	None	None	1
Numbers with…	1	None	None	1
TOTAL:	**263**	**19**	**9**	**291**

USAGE: Hard, hardly
Hardly is not the adverb of **hard**. The adverb of **hard** is **hard**:
   I work hard.
✗ Don't say: I work hardly.
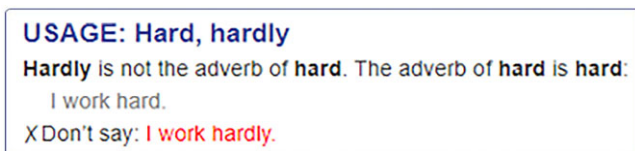
Figure 4: A Type 3 warning note: HARD (adv.)

COLLOCATIONS
VERBS
ADJECTIVES
RESEARCH + NOUN
PHRASES
COMMON ERRORS

Figure 5: Typical headings in a COLLOCATIONS box

COMMON ERRORS
▶ Don't say 'make research'. Say *do research* or *carry out research*.
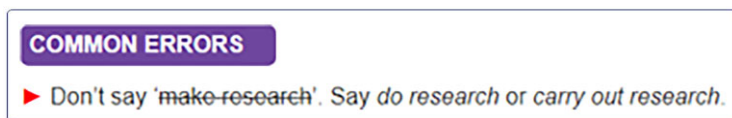
Figure 6: A Type 4 warning note: RESEARCH *(n)*

items, presented in bold, and is marked by a black cross (✗) and a 'Don't say' note followed by an erroneous sentence in red (see Figure 4).

In this study we did not investigate the use of Type 3 warning notes in *LDOCE Online*, because there were only 27 notes of this type for words in *LC9000*, and none of these were for the selected target words in our correction task.

Type 4:
Type 4 warning notes are presented within a framed collocations box with a blue bold heading in capital letters. Below this heading is a list of further headings written in white characters against a purple background indicating collocations with words in various word classes. The final heading on the list is always COMMON ERRORS (see Figure 5).

The Type 4 warning note comes under the COMMON ERRORS heading, and consists of a red triangle (▶) and a 'Don't say' note (see Figure 6). The word or phrase in the common error is crossed out and followed by a 'Say' note with a more appropriate alternative collocation in italics. Type 4 notes are typically positioned before the 'Thesaurus' or 'Examples from the Corpus' parts of the dictionary.

2.2. Research Questions
One step towards systematising the amount and type of typographical enhancement provided in dictionary warning notes would be to discover the formats that are most noticeable

to users, and that lead to the greatest productive success. This study therefore addresses two main research questions:

1. To what extent do the warning notes in *LDOCE Online* help learners of English to correct errors?
2. To what extent does the visual presentation of the warning note affect the success of error correction?

3. Methodology

3.1. Participants

We gathered participants by posting the link on professional network sites, and also by sending it to high school teachers and university instructors who were known to us. These practitioners passed on the link to their own students, or to colleagues who could in turn pass on the link to students or to other colleagues who could identify students who would take part in the study. This cascading process enabled us to attract a relatively large number of participants (332) and gather a relatively large number of answers (1,819). However, collecting participants by this means also meant that they were not known to us personally and we did not have their contact details. Under these circumstances we had no way of placing them in comparable, equal-sized groups which could each take a different version of the task with the words in a different order. Thus, in order to acquire a larger dataset of answers we sacrificed some degree of control. All the participants were learners of English: 279 (84%) were university students and 36 (11%) were high school students. Most (314) were based in Turkey, but a few (17) (5%) were based in Greece, Indonesia and Poland.

3.2. Selection of lexical items and warning notes

Five variables were taken into account when designing questions for our test: the frequency of the target words, the length of the dictionary entry, the position of the warning message in the dictionary entry, the length of the warning message and the format type of the warning message. We wanted to use fairly frequent words in relatively short entries – no more than 50 words - so that participants would not have to scroll very far down the entry to find the relevant warning message. We also wanted relatively short warning messages - 25 words on average in our sample – representing different formatting types in equal numbers. We were aware that position might influence noticing: it is widely accepted that the first sense of a dictionary entry tends to have special salience (Bogaards 1998, Lew et al. 2013, Nesi and Haill 2002, amongst others), although Dziemianko (2014) found that collocations were retrieved most successfully from the bottom of an entry, and Nesi and Tan (2011) suggested that experienced users might take particular notice of the final sense. However, using authentic dictionary data it was not possible to obtain a perfect balance between word frequency, entry length, warning type length and position and warning type representation.

Our target words were selected from the top 1000 words in the *New Academic Vocabulary List* (*AVL*) (Gardner and Davies, 2020). This list is derived from a 120-million-word academic subcorpus of the 425-million-word *Corpus of Contemporary American English*. Three target words were selected for each of the four warning types under investigation, making 12 words in total, five of which were nouns and seven verbs. We believed that this was an appropriate number of items to represent all four warning types without placing too much demand on participants' time and ability to concentrate. All the words

belonged to the high frequency or medium frequency bands in the *Longman Communication 9000*. We rejected high frequency words with long warning notes, such as 'disagree', which has a 115 word Type 2A note. All the *LDOCE Online* warning notes for the target words are provided in Supplementary Online Material.

Table 4 shows the 12 target words in the order in which they were presented in the correction task, together with their frequency, warning note type, (sub)headings, entry position and error type. The chosen order provided a mix of these elements, so that, as far as possible, users were presented with a succession of different types of errors, in different positions in entries containing different types of warning. Only two entries have more than one warning note: the entry for KNOWLEDGE (item 9) which has both Type 2B and Type 4 warnings, and the entry for CONSIST (OF) (item 12) which has two Type 2B subheadings. In both cases only one of the warnings was relevant to the correction task.

Each of the four warning types was presented in entries for three different words. In seven cases the warning message related to the first senses of the word listed in the entry; in one or two cases it was necessary for the respondent to scroll down to find the warning message.

3.3. The Correction Task

Each of the target words was contextualized in a sentence extracted from the British Academic Written English (BAWE) corpus (www.coventry.ac.uk/BAWE) using the GDEX (Good Dictionary Examples) function in *Sketch Engine* (Kilgarriff et al. 2008), which enabled us to identify accessible contexts to illustrate target word meaning. Care was taken to ensure that each sentence was of a similar length.

We then manipulated each sentence so that it required a correction of the type highlighted in the relevant warning note. In some cases we also altered the sentence slightly to make it easier to read. For example, Sentence 1, the original BAWE sentence, was converted to Sentence 2, which matches the warning note for CONSIST: '**Consist** is not used in the passive. ✗Don't say: My family is consisted of four people'.

1. The group <u>consisted of</u> the Operations Director, the Operations Manager and the Project Manager, which allocates resources and monitors progress.
2. The group which allocates resources and monitors progress <u>is consisted of</u> the Operations Director, the Operations Manager and the Project Manager.

The correction task started with the instructions and the example shown in Figure 7.

Participants were presented with each sentence in turn. They were not told which part of the sentence needed to be corrected, but were given the option to access *LDOCE Online* via a hypertext link which indicated the entry where the relevant information would be found (see Figure 7). The Task was deliberately entitled 'Correction with the online *Longman Dictionary*' and did not include any mention of 'warning notes', as we wanted to see whether respondents would notice the warning notes without any prompting. After each correction they were asked whether or not they had clicked on the link we had provided.

The platform used to gather data was an online Google survey. The first part gave some introductory information about the correction task and asked participants about their study status and the country they were studying in. The second part was the main section where the respondents were expected to correct errors using the online dictionary links. Once they

Table 4: Words selected for the study

	Target Word	LC9000 Band	Warning type	Heading/ subheading	Sense Number	Position of warning note
1	DIFFICULTY *n.*	High	TYPE 1	None	Sense 1 of 5 *if you have difficulty doing something, it is difficult for you to do*	After Sense 1
2	PROVIDE *v.*	High	TYPE 2B	**GRAMMAR: Patterns with provide**	Sense 1 of 3 *to give something to someone or make it available to them, because they need it or want it*	After Sense 3
3	CONTRIBUTE *v.*	Medium	TYPE 2A	**Grammar**	Sense 1 of 3 *to give money, help, ideas etc to something that a lot of other people are also involved in*	After Sense 3
4	EXPERIMENT *n.*	High	TYPE 4	**COMMON ERRORS**	Sense 1 of 2 *a scientific test done to find out how something reacts under certain conditions, or to find out if a particular idea is true*	After Sense 2
5	CONTAIN *v.*	High	TYPE 2B	**GRAMMAR: Using the progressive**	Sense 2 of 6 *if a document, book, speech etc contains something, that thing is included in it*	After Sense 1
6	PERSIST *v.*	Medium	TYPE 2A	**Grammar**	Sense 1 of 2 *to continue to do something, although this is difficult, or other people oppose it*	After Sense 1
7	QUALITY *n.*	High	TYPE 4	**COMMON ERRORS**	Sense 5 of 5 *quality of life*	After Sense 5
8	PERFORM *v.*	High	TYPE 1	None	Sense 1 of 3 *to do something to entertain people, for example by acting a play or playing a piece of music*	After Sense 1
9	KNOWLEDGE *n.*	High	TYPE 4	**COMMON ERRORS**	Sense 1 of 2 *the information, skills, and understanding that you have gained through learning or experience*	After Sense 2, at end of long entry
10	LACK *v.*	Medium	TYPE 1	None	Sense 1 of 2 *to not have something that you need, or not have enough of it*	After Sense 1
11	NATURE *n.*	High	TYPE 2A	**Grammar**	Sense 1 of 1 *everything in the physical world that is not controlled by humans, such as wild plants and animals, earth and rocks, and the weather*	After Sense 1
12	CONSIST (OF) *v.*	High	TYPE 2B	**Grammar Comparison**	Sense 1of 1 *to be formed from two or more things or people*	After Sense 1

INSTRUCTIONS:

In each of the following sentences something is wrong. Please copy the part of the sentence which is not quite right, and correct it. You don't need to re-write the whole sentence. Here's an example for you:

0. Example: Psychological treatment should focus in support and encouragement in order to achieve changes in clients' interaction with other people. https://www.ldoceonline.com/dictionary/focus

0. ANSWER: ... focus ON ...

Figure 7: Instructions for the Correction Task

had competed and submitted each correction they were not allowed to return to it, or adjust any previous answer.

3.4. The Pilot study

The data collection instrument was piloted with 15 volunteers. Three were teachers of English, who tended to comment on the suitability of the sentences we had chosen, and 12 were undergraduate learners of English, who tended to comment on the suitability of the task. Some participants tried to correct sentences by changing the word order and/or using alternative words or phrases; the results were grammatically acceptable but it was sometimes difficult to tell the extent to which the guidance in the dictionary warning note had been followed. We altered some of the sentences to reduce the likelihood of corrections of this kind occurring. Some participants thought that they were only supposed to use the dictionary links as a last resort, so we added an extra instruction to click on the link if in any doubt about how to correct the sentence, and we provided more information about the purpose of the task, clarifying that it was intended to investigate online dictionary use for language production, and not as a measure of language proficiency. At this point we also added a third part to the task, for comparative purposes. In this part examples of each type of warning note were presented, together on one page, and participants were asked to choose which type they thought most useful. The final version of the correction task is reproduced in Supplementary Online Material.

On the basis of the pilot study responses we developed a system for coding corrections, shown in Table 5. We discarded all responses produced without clicking on the dictionary links. Of those produced after clicking on the dictionary links, alterations to sentences which used the target words inappropriately and/or changed the meaning of the sentence were counted as unsuccessful (Incorrect), and alterations to sentences which were linguistically appropriate and retained the meaning of the sentence were counted as successful (Correct). A relatively small number of responses evidenced 'uptake' but not 'intake' by reproducing information provided in the warning notes rather than correcting the sentences provided (Noticed but not applied).

The design of the study meant that we could not be absolutely sure whether or not participants had noticed the warning notes, except in the few cases where they reproduced information provided in the warning notes rather than correcting the sentences (Noticed but not applied). Moreover, although participants producing correct answers are likely to have been influenced by the information in the warning notes, they also probably took into

Table 5: Data coding system and examples

Codes	Examples for Item 3: "Army doctors *contribute* for the functioning of an army but are not considered legitimate targets by international law."
Incorrect	• '*…contribute to for…*'
Correct	• '*contribute to the functioning*'
Noticed but not applied	• '*Not contribute for something*'

account information provided in other parts of the dictionary entry, as we would expect in normal dictionary consultation.

3.5. Interrater reliability checks

To establish reliability in the coding of data from the main study, two raters (the first and second authors) independently rated all the corrections made after using the dictionary links provided. The agreement rate between the two raters was 88%. We took several steps to resolve the 12% of cases where the raters disagreed. First we established that human error had resulted in some intra-rater inconsistencies. Then we identified borderline cases where the two raters had reached different conclusions about acceptability. For example, the raters had different views about 'persist in', rather than 'persist in thinking' as a correction for the sentence containing the pattern "*men …. persist to think*". Similarly, the raters had responded differently to cases where participants had provided the correct pattern but had made a mistake with the form of the target word (e.g. '*difficult in adapting…*'; '*can only be gain*'), or used the wrong tense, as in '*do an experiment*' (present tense) to correct '*made an experiment*' (past tense). We agreed to adopt a somewhat lenient policy and accept such instances as correct responses. On the other hand we agreed not to count as correct responses alterations such as *the laws/forces of nature* (to correct *the love of the nature*). In this case the original error (concerning the use of the article) had been corrected, but the respondent has made the change by copying an example from the dictionary entry with a different meaning from that of the sentence in the correction task.

We decided to count as correct those responses containing patterns that remained true to the meaning of the original task sentence, whilst not being the focus of the warning message. For example some respondents wrote '*…provide an independent income for women…*' rather than '*…provide women with an independent income…*'. The Type 2B warning message, "Patterns with provide", gave examples of both structures, but the non-example focussed on the second one ("✗ Don't say: We provide parents information").

4. Results

The total number of participants in the main study was 332. This meant that there were 3,984 answers, as each participant responded to 12 sentences (three for each of the four types of warning message). Of these, 1,846 answers (46%) were produced after clicking on dictionary links. Table 6 shows the proportion of look-ups for each of the three-word sets.

Table 6: Type-based Look-up Counts and Look-up Rates

Type	Total Look-ups	Look-up Rates
TYPE 1	471	48.76%
TYPE 2A	379	39.23%
TYPE 2B	503	52.07%
TYPE 4	493	51.04%
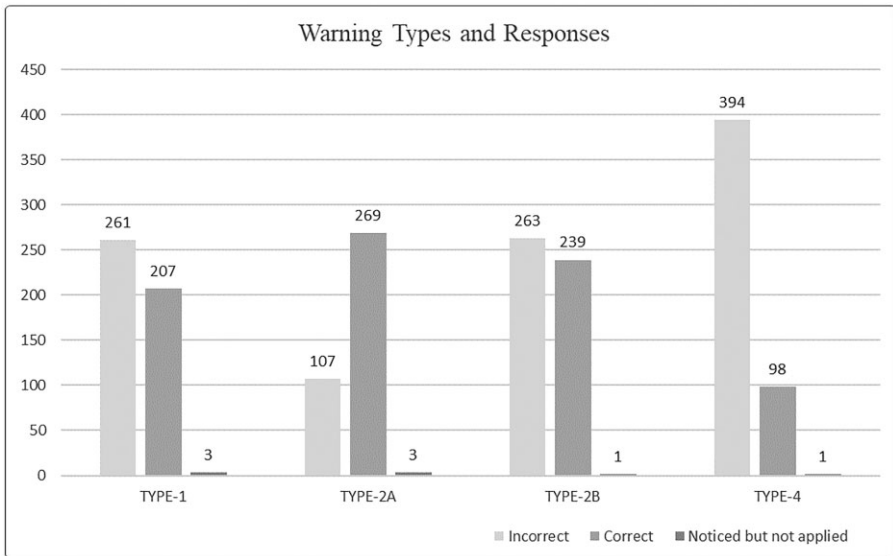
Figure 8: Warning Types and Responses

The numbers were compared to their expected values, using X^2 (3, $N = 1846$) = 20.826. The number of look-ups for Type 2A was significantly lower than its expected value, and the number of look-ups for Type 2B was significantly higher than its expected value. This suggests that participants found it less difficult to correct the sentences containing the Type 2A words, and more difficult to correct the sentences containing the Type 2B words. However these differences were in fact largely due to look up patterns for only one of the words in each three word set. NATURE (Type 2A) had significantly fewer look-ups than expected ($p < .001$), and PROVIDE (Type 2B) had significantly more look-ups than expected ($p < .05$). This could not have been predicted, as both words are in the high frequency band in *LC9000*.

Figure 8 indicates that responses given after exposure to Type 4 warning messages were the least likely to be correct, and responses given after exposure to Type 2A warning messages were the most likely to be correct.

Chi-Square analysis indicated significantly different distributions of correct and incorrect responses across the warning types (χ^2 (3, $N = 1778$) = 240.466, $p < .001$). Table 7 shows that the correct responses after exposure to Type 1 and Type 2B warning messages corresponded to the mean ratio with which the collated values are compared and presented

Table 7: Distribution of Observed Values and Success Rates

	Observed Values		Success Rate
	Incorrect	Correct	
TYPE 1	261	207	44.23%
TYPE 2A	107	269	71.54%
TYPE 2B	263	239	47.61%
TYPE 4	394	98	18.76%

Table 8: Respondents' evaluation of the usefulness of warning note types[*]

Which warning type in the dictionary do you think is most useful?	Frequency	Percentage
Type 1	38	11.45%
Type 2B	120	36.14%
Type 2A	54	16.27%
Type 4	71	21.39%

*These numbers do not add up to 100%, as some respondents showed no clear preference, or preferred other types of note not discussed in this paper.

in the success rate column. However, the correct responses after exposure to Type 2A messages were higher than the mean ratio and the correct responses after exposure to Type 4 messages were lower than the mean ratio.

Table 8 shows the respondents' evaluation of the usefulness of different types of warning note. Here, Type 2B is the clear winner, followed by Type 4, Type 2A and Type 1.

A one-way ANOVA was performed to compare the effect of TYPE (as the independent variable) on Response Rate as the dependent variable indicating the correct response rates for each participant. The findings from the ANOVA test revealed that there was a statistically significant difference in Response Rate between at least two groups ($F(3, 991) = 36.345$, $p<.001$). Tukey's HSD Test for multiple comparisons indicated that the mean value of Response Rate (the dependent variable) was significantly different between TYPE-1 and TYPE-2B ($p <.05$, 95% C.I. = [-.319, .552]), TYPE-1 and TYPE-4 ($p< 0.01$, 95% C.I. = [-.150, .675]), TYPE-2A and TYPE-2B ($p< 0.01$, 95% C.I. = [-.251, .604]), TYPE-2A and TYPE-4 ($p< 0.01$, 95% C.I. = [-.202, .608]), TYPE-2B and TYPE-4 ($p< 0.001$, 95% C.I. = [-.021, .779]).

5.   Discussion

The respondents' order of preference in terms of usefulness, shown in Table 8, matches the results we might have predicted on the basis of the level of typographical enhancement (see Table 1). Type 2B and Type 4 warning messages have the greatest number of typographical features (six), followed by Type 2A (five) and Type 1 (four).

However, when respondents were asked to evaluate usefulness they were shown example warning messages out of context rather than within full dictionary entries, so they

could not see where within entries the messages were placed. For the correction task where participants had access to the complete entries the results were not so clear cut: Type 2A warning notes with five typographical features led to far greater success than Type 4 warning notes with six. Type 2B notes with six typographical features did not lead to much more success than Type 1 notes, which only had four.

There are various possible explanations for this result. Although the Type 4 warnings were more visually prominent than Type 2A and Type 1, they tended to be placed lower down the dictionary entries and were therefore less accessible to participants. Type 1 warning messages were not so visually prominent but always came after Sense 1. Types 2A and 2B are very similar, and although if both types occur within the same entry Type 2A comes before Type 2B, in our study none of the entries included both of these types, so they both occupied similar positions (after Sense 1 on two occasions, and after Sense 3 on one occasion). The competing theories of Cognitive Load (Sweller et al. 1998) and Involvement Load (Hulstijn and Laufer, 2001), would suggest that, because of the combined effect of positioning and visual enhancement, these two types might lead to more noticing (and ultimately retention, although our study did not measure this). Sweller et al. (1998) argue that clearer presentation of information decreases the learner's extraneous cognitive load and therefore increases the chances of learning taking place, while Hulstijn and Laufer (2001) argue that if vocabulary information requires more processing (as, for example, in cases where the presentation is more complex) this leads to greater retention.

6.   Conclusion

As we have seen, a variety of fonts and symbols are used in warning notes in *LDOCE Online* in an attempt to prevent them from going unnoticed in the midst of dictionary content. The use of these various typographical enhancements seems quite unsystematic, and we could not discern the reason for the presence of some enhancements in some warning notes but not in others. The types did not always have an obvious association with one kind of language error rather than another, and indeed in practice it is often difficult to place errors in distinct categories (verb patterns, grammar, usage or collocation). Sometimes the same kind of error is given a different type of warning note in different entries in the dictionary. Warning notes of the kind we have investigated (containing a non-example) have not previously been studied either descriptively or experimentally, and our typographical classification and the finding that the enhancements are not consistently applied are therefore major contributions to the field.

However, our study does have limitations. It was 'naturalistic' in the sense that an authentic online dictionary was used, and participants could complete the survey in their own locations and in their own time, choosing whether or not to consult the dictionary to help them correct the errors in the task. We wanted to observe the effect of the four warning note types on error correction success, and we tried as far as possible to reduce the effect of the confounding variables of target word frequency, dictionary entry length, sense position in the entry, and warning note length and position. In naturally occurring data, however, it proved impossible to create a test where all these variables were very tightly controlled. Moreover although all the target words were relatively frequent, and all the warning messages attached to these words related to common errors in the Longman Learner Corpus, the test items did not all turn out to present the same level of difficulty to our participants. It seems relatively safe to assume that participants were honest about whether they clicked on a dictionary link, and that they only clicked on a dictionary link when they were unsure about how to correct the sentences in the Correction Task. However we were unable to

predict and adjust for the fact that some equally frequent words and patterns proved to be more problematic than others, irrespective of the warning message type. Possible further confounding factors were the fixed order of the questions (as participant behaviour may change over the course of a series of questions), and the impossibility of confirming whether participants had actually noticed the warning messages, even if they said they had clicked on the links. In some cases information that would have helped correct errors was available in other parts of the dictionary entries, outside the warning notes. Thus, although our findings might on balance suggest that borders and headings and more prominent placement might help users notice and apply warning note information, we cannot say with certainty whether the type of warning note was really responsible for the observed differences in correction performance.

Our chosen procedure seemed relatively authentic in that it came closer to replicating the actual consultation behaviour of learners when checking the accuracy of their own written output. The next stage of investigation should, however, build on the findings from this study to test the effectiveness of each typographical feature in isolation. Our participants were recruited across a fairly wide geographical area, via mailing list posts and word of mouth. Most of them were not known to us personally. This made it possible for more people to take part, especially during 2020 lockdown restrictions, but also made it impossible to divide them into comparable groups to take different versions of the Correction Task. In a more controlled type of experiment there could be multiple versions of the task, presenting the target words under different conditions. It seems that the only safe way to control for entry length and sense position, word idiosyncrasies, and the chance that some errors would prove easier to correct than others, would be to create a purpose-built dictionary consisting of a series of purpose-built dictionary entries, controlled for length and sense position but with different types of warning note. The target words could then be rotated across test versions so that each group of participants saw each word in a different sequence, linked to a purpose-built entry showing a different type of warning note. If each warning type was to be linked to three target words, as in the current study, this tight control would entail very many versions of the test. An alternative approach, using eye-tracking technology, would be the best way to check whether participants were actually noticing the warning messages, and the interplay between correction success and the length and direction of their gaze. Both these methods would probably only be feasible with far fewer participants, because of the need for laboratory conditions.

Certainly there is plenty of scope for further research in this area, as warning notes are an increasingly important feature in most learners' dictionaries, and have the potential to greatly improve the effectiveness of dictionary consultation, if they are properly noticed.

Acknowledgements

Supplementary material

All the *LDOCE Online* warning notes for the target words and the correction task are available at *International Journal of Lexicography* online.

References

A. Dictionaries

Delacroix, L. (ed.). 2014. *Longman Dictionary of Contemporary English* (Sixth Edition.) Harlow: Longman. Free online version: https://www.ldoceonline.com/

Mayor, M. (ed.). 2009. *Longman Dictionary of Contemporary English* (Fifth Edition.) Harlow: Longman.

B. Other literature

Bogaards, P. 1998. 'Scanning Long Entries in Learners' Dictionaries.' In **Fontenelle, T., P. Hiligsmann, A. Michiels, A. Moulin and S. Theissen** (eds), *Proceedings of the 8th EURALEX International Congress, EURALEX 2008, Liège, Belgium, August 4-8, 1998.* Liège: Université de Liège, 555–563.

Chan, A. Y. W. 2011. '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.

Chan, A. Y. W. 2012. 'A Comparison Between *COBUILD*, *LDOCE5* and *CALD3*: Efficacy and Effectiveness of the Dictionaries for Language Comprehension and Production' In **Fjeld, R.V. and J. M. Torjusen** (eds), Proceedings of the 15th EURALEX International Congress, EURALEX 2012, Oslo, Norway, August 7-11, 2012. Oslo: Universitetet i Oslo, 606–612.

Craik, F. and R. S. Lockhart. 1972. 'Levels of Processing: A Framework for Memory Research.' *Journal of Verbal Learning and Verbal Behavior* 11.6: 671–684.

Craik, F. and E. Tulving. 1975. 'Depth of Processing and the Retention of Words in Episodic Memory.' *Journal of Experimental Psychology* 104.3: 268–294.

Dai, Y., Z. Wu and H. Xu. 2019. 'The Effect of Types of Dictionary Presentation on the Retention of Metaphorical Collocations: Involvement Load Hypothesis vs. Cognitive Load Theory.' *International Journal of Lexicography* 32.4: 411–431.

Diemand-Yauman, C., D. Oppenheimer and E. Vaughan. 2011. 'Fortune Favors the Bold (and the Italicized): Effects of Disfluency on Educational Outcomes.' *Cognition* 118.1: 111–115.

Dziemianko, A. 2015. 'Colours in Online Dictionaries: A Case of Functional Labels.' *International Journal of Lexicography* 28.1: 27–61.

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.

Firth, J., J. Torous, B. Stubbs, J. Firth, G. Steiner, L. Smith, M. Alvarez-Jimenez, J. Gleeson, D. Vancampfort, C. Armitage and J. Sarris. 2019. 'The "Online Brain": How the Internet May Be Changing Our Cognition.' *World Psychiatry* 18: 119–129.

Gardner, D., and M. Davies. Academic Vocabulary Lists. Accessed on 10 March 2020: https://www.english-corpora.org/coca/ or http://www.academicwords.info

Gouws, R. H. and D. J. Prinsloo. 2010. 'Thinking Out of the Box–Perspectives on the Use of Lexicographic Text Boxes' In **Dykstra, A. and T. Schoonheim** (eds), *Proceedings of the 14th EURALEX International Congress, EURALEX 2010, Leeuwarden/Ljouwert, the Netherlands, July 6-10, 2010.* Leeuwarden/Ljouwert: Fryske Akademy, 501–511.

Hulstijn, J. H. and B. Laufer. 2001. 'Some Empirical Evidence for the Involvement Load Hypothesis in Vocabulary Acquisition.' *Language Learning* 51.3: 539–558.

Kilgarriff, A., M. Husàk, K. McAdam, M. Rundell and P. Rychly. 2008. 'GDEX: Automatically Finding Good Dictionary Examples in a Corpus.' *Proceedings of the 13th EURALEX*

International Congress, EURALEX 2008, Barcelona, Spain, July 15-19, 2008. Barcelona, Universitat Pompeu Fabra, 425–432.

Lee, S. K. and H. T., Huang. 2008. 'Visual Input Enhancement and Grammar Learning: A Meta-analytic Review.' *Studies in Second Language Acquisition* 30.3: 307–331.

Lew, R., M. Grzelak and M. Leszkowicz. 2013. 'How Dictionary Users Choose Senses in Bilingual Dictionary Entries: An Eye-tracking Study.' *Lexikos* 23: 228–254.

Nesi, H. and R. Haill. 2002. 'A Study of Dictionary Use by International Students at a British University.' *International Journal of Lexicography* 15.4: 277–305.

Nesi, H. and K. H. Tan. 2011. 'The Effect of Menus and Signposting on the Speed and Accuracy of Sense Selection.' *International Journal of Lexicography* 24.1: 79–96.

Osada, T., J. Sugimoto, Y. Asada, and Y. Komuro. 2015. 'An Analysis of Longman Dictionary of Contemporary English, Sixth Edition.' *Lexicon* 45: 1–73.

Peters, E. 2012. 'Learning German Formulaic Sequences: the Effect of Two Attention-drawing Techniques.' *The Language Learning Journal* 40:1, 65–79.

Schmidt, R. 2001. Attention. In **Robinson P.** (ed), *Cognition and Second Language Instruction*. New York: Cambridge University Press, 3–32.

Sharpe, P. 1995. 'Electronic Dictionaries with Particular Reference to the Design of an Electronic Bilingual Dictionary for English-speaking Learners of Japanese', *International Journal of Lexicography* 8.1: 39–54.

Sharwood Smith, M. 1993. 'Input Enhancement in Instructed SLA: Theoretical Bases.' *Studies in Second Language Acquisition* 15.2: 165–179.

Stirling, J. 2005. 'The Portable Electronic Dictionary - Faithful Friend or Faceless Foe?' *Modern English Teacher* 14.3: 64–72.

Sweller, J., J. V. Merrienboer and F. Paas. 1998. 'Cognitive Architecture and Instructional Design.' *Educational Psychology Review* 10.3: 251–296.

Taylor, A. and A. Chan. 1994. 'Pocket Electronic Dictionaries and their Use' In **Martin, W., W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg and P. Vossen** (eds), *Proceedings of the 6th EURALEX International Congress, EURALEX 1994, Amsterdam, the Netherlands, August 30-Sept 3, 1994*. Amsterdam: Vrije Universiteit Amsterdam, 598–605.

Zhang, P. 2004. 'Is the Electronic Dictionary Your Faithful Friend?' *CELEA Journal* 27.2: 23–28.