CHAPTER FOUR

NAVIGATING DICTIONARY SPACE:
THE FINDABILITY OF ENGLISH COLLOCATIONS
IN A GENERAL LEARNER’S DICTIONARY
(LDOCE4) AND SPECIAL-PURPOSE
DICTIONARY OF COLLOCATIONS (OCD)

ROBERT LEW AND MAGDALENA RADŁOWSKA

1. Introduction

Accurate selection of collocations is one of the most troublesome, and most elusive, areas of lexical difficulty for foreign learners of English. Modern lexicography can offer assistance in this regard, as a growing degree of attention is paid to word combinations. This is no doubt partially due to the corpus revolution in lexicography, as modern corpus methods can provide fairly objective evidence on the cooccurrence of lexical items. The other important factor is the growing awareness in lexicography of the importance of the dictionary users, their lexical needs and reference skills, which is reflected in the improved design of dictionaries.

The present study undertakes to assess the success with which intermediate pre-university Polish learners of English are able to locate English collocations in two learners’ dictionaries, representing two different dictionary types. The first of these is a general dictionary for learners, Longman Dictionary of Contemporary English, fourth edition (LDOCE). The second dictionary is a special-purpose dictionary dealing specifically with collocations, the Oxford Collocations Dictionary (OCD). The collocations are classified into several types for the findability test.

Numerous studies point to collocations being a highly problematic area for language learners. For example, Howarth (1996) noted collocational errors as an error type persistent in written compositions by advanced students of English. Channell (1980) asked advanced students to select
word combinations within a collocational grid. They gave many inappropriate answers, even though the words were very familiar to these students. Howarth (1998: 42) observes that

learners’ difficulties lie chiefly in [the area of restricted collocations], since idioms and free collocations are, phraseologically, largely unproblematic. The greatest challenge lies in differentiating between combinations that are free and those which are somehow limited in substitutability.

He further adds that it is not just the learners but also many of the teachers who need substantial guidance in the area of collocation. There may be several reasons why collocation presents so much difficulty to language learners, but one nontrivial consideration is that the nature of collocation is probabilistic, statistical: no hard-and-fast rules can be given which learners could memorize or drill.

Collocation certainly presents a serious challenge to language learners, but it is indispensable if the language is to approximate the native norm, be it in spoken or written production. Therefore, dictionaries should provide learners with a solid coverage of word combinations.

The present study undertakes to assess the success with which intermediate pre-university Polish learners of English are able to locate English collocations in two learners’ dictionaries, representing two different dictionary types. The first of these is a general dictionary for learners, *Longman Dictionary of Contemporary English*, (LDOCE), in its fourth edition. The second dictionary is a special-purpose dictionary dealing specifically with collocations, the *Oxford Collocations Dictionary* (OCD).

2. Dictionaries and collocation

2.1. Collocational patterns

In this study we focus on lexical, as opposed to grammatical, collocation, which in most common formulations are two-word pairings consisting of a N/Adj/Verb/Adv *base* and a *collocator*. The distinction is attributed to Hausmann (1985), who proposed to separate the combination of lexical items into *Basis* and *Kollokator*. The base is assumed to be the dominant word which carries most of the whole collocation’s meaning, whereas the collocator serves a subordinate function and is base-dependent. For instance, in *draw attention*, the noun *attention* would be the base; the verb *draw*, the collocator.
The popular BBI dictionary (Benson et al. 1998) identifies six specific types of lexical collocation:

1. V+N: fly a kite
2. Adj+N: weak tea
3. N+V: blood circulates
4. N1 of N2: a bouquet of flowers
5. Adv+Adj: strictly accurate
6. V+Adv: apologize humbly

(BBI: xxx)

The newer OCD includes all the above patterns but adds three more:

1. N+N: a light source
2. V+V: be free to choose
3. V+Adj: make/keep/declare sth safe

(OCD 2002: ix)

For the purpose of the present study, we will follow the above classification, that is the nine patterns of lexical collocations recognized in OCD.

### 2.2. Where to place collocations in dictionaries

Based on his distinction into base and collocator, Hausmann proposed that the strategy of placing collocations in a dictionary should reflect its primary function: if a dictionary is aimed at helping with encoding, collocations should be placed at the base. In contrast, in a dictionary designed primarily for decoding, collocations should be entered at the collocator. Any collocations located in contradiction of the above principle were termed hidden by Hausmann. His approach has had a considerable impact on lexicographic practice. However, empirical research by Bogaards (1999) does not fully corroborate Hausmann’s neat distinction into base and collocator, which has also come under theoretical criticism (e.g. Siepmann 2005).

### 2.3. Coverage and placement of lexical collocations in OCD and LDOCE

OCD is a dictionary specifically focusing on collocation. It claims (back cover, also Walker 2009) to cover some 150,000 collocations, grouped at approximately 9,000 entry words: nouns, verbs, and adjectives.
Most of the collocations covered were retrieved from the 100-million-word British National Corpus (OCD: viii-ix). Within entries, collocations are grouped according to sense (if any), then syntactic pattern, and within each pattern collocations may be further clustered based on semantic relatedness. Some collocations (though by no means all: OCD’s back cover claims 50,000 examples) are supplied with example sentences. Sense discriminators are given in cases where the editors saw a need to indicate differences between particular senses placed at the same headword.

In general, OCD adopts Hausmann’s approach to the placement of lexical collocations. As it is primarily a dictionary for production (Walker 2009), collocations are on the whole arranged by Hausmann’s base. One consequence of this is that the verb and adjective entries in OCD do not include any noun items as collocates (Klotz 2003).

LDOCE, not being a collocation-specific dictionary, deals with general language and provides information about meaning, spelling, pronunciation, grammar, etc. Collocation is but one of many aspects treated in the dictionary. The main source of words for LDOCE (in its 4th edition) had been the then 300-million-word Longman Corpus Network (LDOCE: x). In the dictionary, collocations are illustrated in coloured boxes (mainly blue) and they are incorporated into examples drawn from the corpus. Additionally, many collocations are provided with glosses, e.g. welcome news (= good news) (LDOCE: 1105), and this, as we shall see, appears to be significant when it comes to aiding comprehension. Collocations tend to be placed at the base, but there are obviously many hidden collocations. The dictionary does not specify the number of collocations covered, but gives a total for all word combinations as 220,000 (LDOCE back cover).

Unlike OCD, LDOCE is not exclusively a production dictionary: it is also intended to offer assistance in reception, and this has consequences for the placement of collocations. For example, if we look at the headword draw verb, we will see that LDOCE provides a variety of combinations for this entry, such as: draw attention, draw blood etc., yet it does not include them at the noun base (LDOCE 2003: 474).
3. The study

3.1. Aim, design and procedure

The aim of the study was to see which of the two dictionaries, OCD or LDOCE, would serve Polish Secondary School learners better on a task of supplying missing collocations in sentences.

Eighteen students participated, all attending the 1st Secondary School in Goleniów. They were all native speakers of Polish, 18 and 19 years old, soon to take their secondary school leaving exam (*matura*). They attended English classes five times a week for 45 minutes.

Students were requested to supply missing words in 13 gapped sentences in a way that would fit the context. The task was done with the assistance of dictionary booklets, consisting of dictionary entries relevant for the test items. There were two versions of the booklets: nine subjects were supplied with entries from the OCD, the other nine from LDOCE. To explain how they should go about the task, subjects were provided with one example that contained two sentences. The first sentence was formulated in Polish, and the second one was its partial English translation with a gap to be filled in (see Appendix 1). The test proper consisted of 13 items representing 9 different types of lexical collocations: V+N, Adj+N, N+V, N+N, Adv+Adj, V+Adj, V+V, V+Adv, N1 of N2 (see section 0). We made sure that all of the tested collocations were covered by both OCD and LDOCE.

Students were supplied with a brief instruction on how to use each dictionary. Then, the two versions of dictionary booklets (one with OCD entries, the other with LDOCE entries) were distributed. The time assigned to carry out the task was limited to 25 minutes. The session was held on January 31, 2008.

During the test quite a few participants appeared to be confused by some items, and finding the correct answer posed a problem for them, despite their relatively high level. This underscores the difficulty of collocation for learners.

At the scoring stage, one point was granted for each correctly given collocation, bringing the maximum total score to 13. Acceptable item combinations are given in the Answer Sheet (see Appendix 2).
3.2. Results

3.2.1. Overall success rates

In the OCD group, the maximum individual score was 12 points out of 13 (92%), the lowest score was 3 (23%), and the group mean was 6.33 (49%). In the LDOCE group, two subjects scored 100% (13 points), while the minimum score was 1 (8%). The overall group mean was 8.89 (68%). Compared with OCD users, LDOCE users scored higher by 2.56 points, a difference of 39% (or 19 percentage points). In order to assess the difference, a 1-Way ANOVA was computed with the absolute score as the dependent variable and dictionary as categorical predictor (with two levels: OCD and LDOCE). The difference turned out not to be statistically significant ($F_{(1,16)}=2.39$, $p=0.14$, n.s.). Thus, we cannot claim with confidence that LDOCE is generally more helpful. Nevertheless, given the small sample size and the promising difference between the means, there is a real possibility that if a larger sample of subjects had been investigated, the difference would have reached significance. To assess this likelihood, we conducted a power analysis for this ANOVA (the power of a test reflects the probability of detecting a real population difference from a sample). At alpha=0.05, observed power of our test turned out to be a low 0.31, no doubt due to a small sample size and large in-group variation. Clearly, then, further research should be carried out on a larger sample.

3.2.2. Success rates by collocation pattern

Figure 1 below gives the success rates broken down by collocation type and dictionary. Out of the nine collocation patterns, LDOCE scored higher for seven types, there was a tie for the Adv+Adj type, and OCD outperformed LDOCE in the one V+V pattern. Let us now examine some of these patterns.
3.3. A closer look at two items

Within the limited space available here, we shall only examine two specific items: one where there was a clear difference between OCD and LDOCE, and one where the two dictionaries performed similarly.

Throw light on sth (V+N)

None of the OCD users were able to provide an acceptable answer for the following, metaphorical, V+N collocation:

1. Odkrycia te RZUCAJ światło na kwestię opracowania szczepionki przeciw AIDS.

These discoveries ............... light on the development of an AIDS vaccine.

Below we give partial entries for the noun light from OCD and LDOCE.

<table>
<thead>
<tr>
<th>OCD</th>
<th>LDOCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>VERB + LIGHT have Have you got enough light for reading?</td>
<td>11 throw / shed / cast light on sth to provide new information that makes a difficult subject or problem easier to understand: Melanie was able to shed some light on the situation. These discoveries may throw new light on the origins of the universe.</td>
</tr>
<tr>
<td>provide, shed light emitted by a star</td>
<td></td>
</tr>
<tr>
<td>sensitive to</td>
<td></td>
</tr>
<tr>
<td>generate, produce</td>
<td></td>
</tr>
<tr>
<td>cast, emit, give (out),</td>
<td></td>
</tr>
<tr>
<td>provide, shed</td>
<td></td>
</tr>
<tr>
<td>be bathed in</td>
<td></td>
</tr>
<tr>
<td>be</td>
<td></td>
</tr>
</tbody>
</table>

The complete OCD entry for light is actually very long, with a large number of potential items. However, it was not the length that seemed to cause our subjects greatest trouble. They appeared to have had no problems getting to the relevant section (VERB + LIGHT), but—once there—they often chose an inappropriate verb, all too often going for give, perhaps because it is a very familiar word. Note that give is presented in the same collocational string in the OCD article as the correct collocates (shed or cast), which implies that participants were looking for the appropriate collocator (VERB), but they did not get it right all the way.

In contrast, LDOCE users had much better success, with 5 correct collocations out of 9. Here, LDOCE provides a brief definition and illustrates the possible combinations in the examples, which apparently
went a long way towards helping our subjects, and made a positive difference relative to OCD.

**Vitally important (Adv+Adj)**

For this item, both dictionaries scored success rates of 78%, with seven out of nine subjects in each group giving correct answers. Here is the item as it appeared in the test:

8. Zawód prawnika jest dla niego NIEZWIĘKŁO ważny. * (do not use VERY or MOST)

A career in law is …………… important for him.

Such a high success rate implies that the dictionaries must have done something right in presenting the data. Let us, then, take a look at the two entries:

<table>
<thead>
<tr>
<th><strong>OCD</strong></th>
<th><strong>LDOCE</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>important adj.</td>
<td>important S1 W1 /…/ adj</td>
</tr>
<tr>
<td>VERBS be, seem</td>
<td>become</td>
</tr>
<tr>
<td>make sth</td>
<td>This is what makes our work so important.</td>
</tr>
<tr>
<td>ADV. critically, crucially, enormously, especially, extremely, hugely, most, particularly, really, terribly, very, vitally</td>
<td>These ideas are considered enormously important.</td>
</tr>
<tr>
<td>increasingly</td>
<td>doubly</td>
</tr>
<tr>
<td>fairly, quite</td>
<td>equally</td>
</tr>
<tr>
<td>These two factors are equally important.</td>
<td></td>
</tr>
<tr>
<td>internationally, regionally</td>
<td>an internationally important site for these rare birds</td>
</tr>
<tr>
<td>economically, functionally, historically, politically, strategically</td>
<td>historically</td>
</tr>
<tr>
<td>historically important buildings</td>
<td></td>
</tr>
</tbody>
</table>
It may at first be somewhat puzzling that subjects in the two groups performed similarly despite the large number of (undiscriminated!) collocates in the OCD. What seems to be key is that in this case most items given under the ADV section are actually acceptable, so semantic distinctions between the collocates are not that important. We might conclude, then, that the success of dictionary users with this item is not so much due to the satisfactory lexicographic treatment, but rather to the fact that this item was not all that difficult to get right.

3.4. What makes a successful entry?

For lack of space, we only discuss a couple of items above, but we have examined closely all of the entries tested, looking at variation in length, item grouping, sense discrimination and typography in search of clues to successful presentation of collocation. We find that most problems are likely to have stemmed from lack of comprehension: too often, subjects did not understand the words, especially those which are more sophisticated and less frequent. Example sentences can be very helpful, but in neither dictionary are they provided for every single combination, probably for reasons of space.

What makes LDOCE different is its systematic attempt to gloss the less obvious collocates when they are presented. These short definitions are supplied in simple vocabulary that apparently does a good job of conveying meaning and discriminating between various collocates. The definitions are further supported with examples.

In contrast, OCD has much less semantic indication of meaning, and fewer examples. This, plus a large supply of combinations, can lead to confusion. We thus conclude that the key to success lies in semantic information and natural context being presented in a way that is accessible and adapted to language learners’ needs. This may be beneficial in decoding and encoding alike.

4. Conclusion

Our results show that even fairly advanced learners had experienced serious problems with locating and selecting appropriate collocations. This suggests that learners are not sufficiently aware of the issues of collocational word pairings, and this is something they should be made aware of.

The differences in performance noted in our study can most likely be attributed to two factors. Firstly, semantic explanation, when present,
provides assistance in assuring that the right collocational item has been found. Secondly, example sentences and phrases can be used as a check that appropriate collocations have been located. These two elements are missing from the special-purpose dictionary of collocations, and it appears that this lexicographic decision negatively affects the findability of items.

Our conclusions coincide quite closely with those in Alonso-Ramos (2008), who investigated 25 learners of Spanish as a second language.

References


**Dictionaries**


Appendices

Appendix 1. Collocation supply test

Z pomocą słownika, wstaw brakujące słowa tak, by naturalnie łączyły się z resztą słów. W niektórych przykładach możliwych jest kilka odpowiedzi. Wybierz jedną z nich.
Przykład:
Mary zamierza ZŁOŻYĆ wizytę swoim starym znajomym.
Mary is going to .....pay.....a visit to her old friends.

1. Odkrycia te RZUCIĄ światło na kwestię opracowania szczepionki przeciw AIDS.
These discoveries ............... light on the development of an AIDS vaccine.

They were busy, so I left. I didn’t want to ............... my welcome.

3. Nie wolno ci przekroczyć granicy bez AKTUALNEGO paszportu.
You are not allowed to cross the border without a ............... passport.

4. Niektóre DRZEMIAŁĘ wulkany przejawiały ostatnio oznaki aktywności.
Some ............... volcanoes have recently shown signs of eruption.

5. Samolot STARTUJE o 11:00 po południu.
The plane ............... at 11:00 p.m.

Maggie collects old newspaper ............... about motorcycles.

7. Wszyscy wiedzą, że Lucy jest inteligentną i WYSOCE elokwentną kobietą.
* (do not use VERY)
Everybody knows that Lucy is an intelligent and ............... articulate woman.

8. Zawód prawnika jest dla niego NIEZWYKŁY ważny. * (do not use VERY or MOST)
A career in law is ............... important for him.

9. Trzy miesiące temu ława przysięgłych UZNAŁA go winnym morderstwa młodego chłopca.
Three months ago the jury ............... him guilty of the murder of a young boy.

10. Jackowi będzie miło słyszeć, że wygrał bilet na koncert swojego ulubionego zespołu rockowego.
Jack will be ............... to hear that he has won a ticket for the concert of his favourite rock group.

11. Spaliśmy DOBRZE, bo łóżko było naprawdę wygodne. * (do not use WELL)
We slept ............... because the bed was really comfortable.

12. Wczoraj się pokłócił. GWALTOWNIE gestykulując, Eddie starał się wytłumaczyć swoje zachowanie.
They had a quarrel yesterday. Gesticulating ............... , Eddie was trying to explain his behaviour.

13. Wciąż jest PROMYK nadziei, że porwane dziecko odnajdzie się żywe.
There is still a ............... of hope that the kidnapped child will be found alive.
Appendix 2. Answers accepted

1. **shed / throw / cast** light on sth
2. **outstay / overstay** your welcome
3. **valid** passport
4. **dormant** volcanoes
5. plane **takes off**
6. newspaper **clippings / cuttings**
7. **highly** articulate
8. **vital / critically / crucially / enormously / especially / extremely / hugely / particularly / really / terribly** important
9. **found / deemed** him guilty
10. be pleased to / be delighted to / be glad to / be gratified to / hear
11. slept **soundly / properly**
12. gesticulating **wildly**
13. **ray / glimmer / spark / flicker** of hope