An Analysis of Three Collocations Dictionaries for Learners of English

Ishii, Yasutake
成城大学

Uchida, Satoru
東京外国語大学

Hayashi, Kyoko
成城大学

Kobayashi, Yuichiro
立命館大学

出版情報：Lexicon. 44, pp.24-53, 2014-10-14. 岩崎研究会
バージョン：
権利関係：

An Analysis of Three Collocations Dictionaries for Learners of English

Yasutake Ishii　　Satoru Uchida
Kyoko Hayashi　　Yuichiro Kobayashi

1. Introduction

This paper is a brief critical review of three collocations dictionaries for learners of English with special emphases on the differences in their headwords, collocates and microstructures, as well as how they are perceived by their users.

A "collocation" of a given word refers to its most typical word combinations, and those typical words the given word combines with are called "collocates"[1]. The mastery of collocations is known to be essential for sounding natural in any language.

The following three dictionaries, given in the order of publication, are reviewed in this paper.

- *Oxford Collocations Dictionary*, Second Edition (2009) ("OCD2" henceforth): The number of pages for the A-Z body part is 958. A CD-ROM accompanies each copy of the print dictionary, and some entries are offered only on the CD-ROM version. On the CD-ROM, in addition to regular searching for headwords, we can also search for collocates given under other headwords, which can be of benefit to users. The print dictionary states about the CD-ROM that "30,000-word index makes searching easy" (back cover), although the number of index keys is 22,864[2] according to our counting. The first edition of this dictionary came out in 2002 without a CD-ROM, and its electronic version was released as part of a standalone CD-ROM, *Oxford Phrasebuilder Genie* in 2003.
- *Macmillan Collocations Dictionary* (2010) ("MCD" hereafter): It has 911 pages in the body. There is no electronic version of this dictionary.
- *Longman Collocations Dictionary and Thesaurus* (2013) ("LCD" below): The main body has 1,447 pages. There is no CD/DVD-ROM version of this dictionary, but a one-year subscription to the online version is provided to those who have bought a copy. Some additional entries, collocates and examples are available on the online version.

According to the publishers' catalogues (Oxford University Press 2013; Macmillan Education 2013; Pearson ELT 2013), the target level of users of all three dictionaries is B2-C2 on the CEFR, Common European Framework of Reference for Languages: Learning, Teaching, Assessment, which is a framework that describes the proficiency level of a learner's foreign language ability, wherein the levels are divided into six stages from A1, beginner, to C2, the most advanced.

This paper consists of four sections excluding this introduction and the conclusion in Section 6. Section 2 compares the headwords in the three dictionaries. Section 3 focuses on the microstructures of each dictionary. Section 4 highlights the differences in collocates across the three dictionaries based on a sampling study. Section 5 deals with the users' perspective by showing the results of our user survey.

(Section 1 by Ishii)

2. Headwords[1)]

This section looks at the headwords in each dictionary, and compares the items across the three dictionaries. For this purpose, we made a complete list of headwords for each dictionary. We made the lists for both print and electronic versions except for MCD, and they include those items that are present either in the print edition or in the electronic version or in both. The following numbers and discussions are based on the lists we have created.

2.1. Number of headwords in each dictionary

The number of headwords in OCD2 is 8,418. Empty headwords for cross reference are excluded from this number, and words which have two or more parts of speech are counted separately; for example, **abuse**[2)] (n)[3)] and **abuse** (v) are counted as two items, which is how all three dictionaries give their entries. The print edition of OCD2 states that it includes "[c]ollocations for 9,000 nouns, verbs and adjectives" (back cover), which could be judged to be dishonest according to our count[4)]. 356 out of 8,418 items are given exclusively in the electronic version.

MCD has 4,305 headwords. There are no empty headwords in this dictionary, so this number is the actual total size. The dictionary states that it has "collocations for over 4,500 carefully-selected key words" (back cover),

so this dictionary can also be criticized for exaggerating the count assuming that our count is correct.

LCD presents 3,819 headwords which include 36 items available only in the electronic version that do not give collocates but thesauri and/or other notes. 741 empty headwords for cross reference are excluded from this number. 550 (besides the above 36) out of 3,819 items are given only in the electronic version. We were not able to find any mention of the number of headwords neither in the print dictionary nor in the electronic version.

The breakdown of the headwords in terms of parts of speech is given in Table 2.1. All three dictionaries share a similar tendency in terms of the proportion of parts of speech; about 60% of the headwords are nouns while adjectives and verbs share roughly 20% each.

Table 2.1
Breakdown of Headwords

	OCD2[5]	MCD	LCD[7]
Adjectives	1,486 (17.7%)	997 (23.2%)	680 (17.8%)
Nouns	5,443 (64.7%)	2,406 (55.9%)	2,483 (65.0%)
Verbs	1,489 (17.7%)	902[6] (21.0%)	573 (15.0%)
Adverbs	—	—	49 (1.3%)
(Total)	8,418	4,305	3,819[8]

In terms of the numbers of headwords, OCD2 can be praised for its relatively wide coverage; users are more likely to find the headwords they want to look up in this dictionary.

2.2. Comparison of the coverage

The number of all different types of headwords covered in at least one dictionary is 9,168. 26.3% of the items, namely 2,409 headwords, are covered in all three dictionaries, while 27.8% or 2,549 words are included in two and 45.9%, i.e., 4,210 items, are present only in one dictionary. The more detailed breakdown of these numbers in terms of parts of speech is laid out in Table 2.2. About 70% (1,676 out of 2,409) of the headwords covered in all three dictionaries are nouns, which conforms to the expectation that we look up nouns more often than other parts of speech in collocations dictionaries.

Table 2.2

Breakdown of All Different Types of Headwords Covered in Three/Two/One Dictionaries[9)]

	3	2	1
Adjectives	391	579	833
Nouns	1,676	1,445	2,405
Verbs[10)]	341	525	891
(Total)	2,409	2,549	4,210

It is natural to expect that the widely covered headwords are frequently used words. In order to confirm whether this is the case, we extracted the first ten headwords in alphabetical order from each group of words that are covered in three/two/one dictionaries in our list, and checked their raw frequency in the British National Corpus (BNC)[11)]. The result is shown in Table 2.3. In general, those headwords that are covered in all three dictionaries are observed more frequently than those items covered in two or one dictionary, with a few exceptions such as **able** (adj) and **about** (prep).

Table 2.3

First Ten Items Covered in Three/Two/One Dictionaries and Their Raw Frequency in the BNC

Covered in 3 Dictionaries		Covered in 2 Dictionaries		Covered in 1 Dictionary	
abandon (v)	4,332	**abbreviation** (n)	229	**abashed** (adj)	44
ability (n)	10,378	**abnormal** (adj)	801	**abhorrent** (adj)	66
abortion (n)	1,488	**abolish** (v)	1,864	**ablaze** (adj)	162
absence (n)	5,859	**absent** (adj)	1,451	**able** (adj)	29,657
abstract (adj)	1,860	**absorb** (v)	2,619	**abode** (n)	250
abuse (n)	3,664	**abstract** (n)	370	**about** (prep)	146,934
academic (adj)	4,612	**absurd** (adj)	926	**abscess** (n)	99
accent (n)	1,764	**abuse** (v)	1,206	**absorbed** (adj)	92
accept (v)	19,811	**abusive** (adj)	277	**abundance** (n)	665
acceptable (adj)	3,608	**accelerate** (v)	1,064	**academic** (n)	965

Now let us see how these headwords are covered in each dictionary. Table 2.4 summarizes the relationship between the coverage and the number of headwords in each dictionary. Although MCD and LCD fall behind OCD2

in terms of the numbers of headwords, they have relatively few headwords that are unique to only one dictionary. Considering the fact that MCD and LCD give at least one example for each collocation or each set of synonymous or related collocates while OCD2 does not (see 3.3.2 for more details), we can say that MCD and LCD focus on frequent headwords and try to give as many examples as possible. Users are likely to find the headword whose collocates they want to know as long as the headword is frequent enough, while in other cases for more infrequent headwords it is plausible that only OCD2 can offer them a solution.

Table 2.4
Relationship between the Coverage and the Number of Headwords in Each Dictionary

	OCD2	MCD	LCD
3	2,409 (28.6%)	2,409 (56.0%)	2,409 (63.1%)
2	2,456 (29.2%)	1,501 (34.9%)	1,141 (29.9%)
1	3,547 (42.1%)	395 (9.2%)	268 (7.0%)
(Total)	8,418	4,305	3,819

While there seems to be no special characteristic found in those headwords covered only in OCD2, the items included only in MCD or in LCD have some features that are worth mentioning. The first 10 items of those headwords covered only in MCD in alphabetical order are as follows: **academic** (n), **accepted** (adj), **account for** (phr vb), **add to** (phr vb), **added** (adj), **adhere to** (phr vb), **adverse** (adj), **advisory** (adj), **allege** (v) and **alleged** (adj). We will soon notice the three phrasal verbs here. The inclusion of phrasal verbs as headwords is a contrastive feature of this dictionary. The items extracted in the same procedure from the LCD are as follows: **about** (prep), **actress** (n), **adult** (adj), **affectionate** (adj), **after** (prep, conj, adv), **again** (adv), **alarm clock** (n), **allergic** (adj), **allow** (v) and **almost** (adv). Prepositions and adverbs are included as headwords only in this dictionary[12)], which is a noticeable feature of this dictionary.

(Section 2 by Ishii)

3. Microstructures

3.1. Introduction

This section examines the microstructures of OCD2, MCD and LCD. The structure of entries in these dictionaries is specially devised to list collocates. The basic framework can be described as follows:

headword 1 (see 3.2)
  sense 1 (see 3.2)
    collocational pattern 1 (see 3.3.1)
      collocates (see 3.3.2)
        examples (see 3.4)
    collocational pattern 2
…
  sense 2
    collocational pattern 1
      collocates
        examples
…
headword 2
…

Fig. 3.1, part of an entry taken from the CD-ROM version of OCD2, is a typical example of how each entry looks.

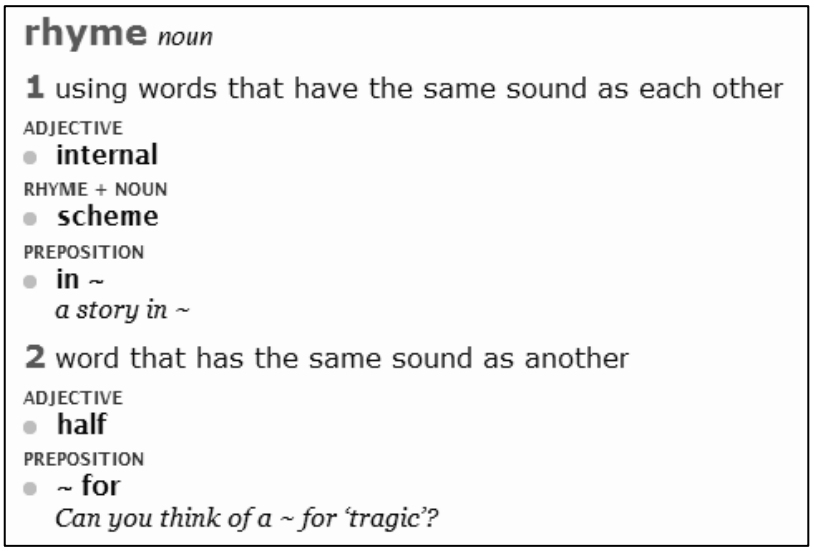

Fig. 3.1. Entry of **rhyme** (n) on the CD-ROM version of OCD2.

We will begin by looking at the framework of entry structures (3.2), and then discuss how collocational patterns and collocates are presented in the entry (3.3) together with examples (3.4) and labels (3.5).

3.2. The basic structure of entries

Let us start by looking at the main framework of entries, that is, headwords and meaning groups. OCD2, MCD and LCD employ the same policy for splitting headwords; they separate entries based on word classes. For example, **review** (n) and **review** (v) are listed as separate entries. Unlike ordinary learner dictionaries, pronunciations of headwords are not presented in the target dictionaries[1)].

Within each entry, headwords with more than one sense are divided into different parts. For example, as can be seen in Fig. 3.1, each sense of **rhyme** (n) ("1 using words that have the same sound as each other" and "2 word that has the same sound as another") is treated separately. In such cases, short definitions of each sense are given so that users can understand which sense of the word goes with the collocates listed under the sense.

There are some differences as to the policy for the definitions. Unlike MCD and LCD, OCD2 does not provide a definition when entries have only one sense, which can be problematic. Take **compatible** (adj) for example, which has the following three senses according to OALD8:

1. (of machines, especially computers) able to be used together
2. (of ideas, methods or things) able to exist or be used together without causing problems
3. if two people are compatible, they can have a good relationship because they have similar ideas, interests, etc.

At the entry of **compatible**, OCD2 lists collocates such as *highly, very, entirely, fully, perfectly* and *mutually*. However, users may wonder which of the above senses can be used with these collocates without the specification of the sense of the headword.

LCD clarifies its policy for ordering each sense of polysemous words as follows: "Meanings of the word are listed in frequency order" (iv). However, OCD2 and MCD do not clearly state their policies in this respect, which could cause user-unfriendliness.

3.3. Collocates

3.3.1. Collocational patterns

Collocates are categorized into several groups according to their syntactic patterns. The basic patterns in each dictionary are as follows[2]:

OCD2

verbs' collocates: adv, v + V, prep, phrases

nouns' collocates: adj, n of N, v + N, N + v, N + n, prep, phrases

adjectives' collocates: v, adv, prep, phrases

MCD[3]

verbs' collocates: adv + V, V + n, and/or

nouns' collocates: adj + N, n + N, v + N, N+ v, N + n, and/or

adjectives' collocates: adv + ADJ, ADJ + n, v + ADJ, and/or

LCD[4]

verbs' collocates: adv, prep, phrases, n

nouns' collocates: adj, adj/n + N, n + N, N + n, v + N, N + v, v, prep, prep/adv, phrases

adjectives' collocates: n, ADJ + n, n + ADJ, adv, v, prep, phrases

While OCD2 and MCD list collocational patterns in the same order in every entry,LCD seems to show collocational patterns in the order of frequency. This results in the following differences in the order of collocational patterns within LCD (print edition):

charter (n): v, phrases, prep

chat (n): adj, v, prep

chocolate (n): phrases, adj/n + **chocolate**, **chocolate** + n

This shows, for example, that "adj + **chat**" is more frequent than "v + **chat**" and "**chat** + prep." Also, we can tell that the "phrases" category is the most frequent for **chocolate**, which lists *a bar of chocolate*, *a piece of chocolate* and *a box of chocolates*.

Let us now look closely at each pattern. First, a note should be made about pre-modifying categories. OCD2 includes many nouns with adjectives because such nouns "function like adjectives" (x). MCD distinguishes nouns

and adjectives that come before nouns. LCD employs a moderate policy: in some cases nouns and adjectives are separate while in other cases they are treated together[5]. The differences can be seen in the entries of **review** (n) as the following excerpts indicate:

review (n) (sense 2) in OCD2
adj: *good*, *bad*, *book*, *film*
review (n) (sense 2) in MCD
adj + N: *good*, *bad*
n + N: *book*, *film*
review (n) (sense 1) in LCD
adj/n + **review**: *a thorough/comprehensive/full review*, *peer review*
review (n) (sense 2) in LCD
adj: *a good/bad review*
n + **review**: *a film/movie review*

We can say that the policy employed in OCD2 is practical, since there are some cases where such distinction is difficult, which might be a motivation for LCD's ambiguous treatment. It should be noted that there seems to be confusion in MCD's treatment as *management* ***committee*** and *cotton* ***curtain*** are treated under the label of "adj + N," though they are more likely to be categorized as the "n + N" pattern.

As Coffey (2010: 334-335) points out, it is one of MCD's characteristic features that it includes noun collocates in verb and adjective entries. LCD also gives noun collocates in its verb and adjective entries. Unlike these two dictionaries, OCD2 excludes noun collocates from verb and adjective entries based on the assumption that a user would not look up noun phrases there[6]. This may hold true for many users, but this can also lead to a serious disadvantage as a collocations dictionary. To take **injure** (v) 'to harm physically' for example, OCD2 (sense 1) lists only adverbs such as *badly, seriously* and *severely*. MCD, in contrast, includes not only adverbs such as *badly* and *critically* (adv + V) but also nouns that indicate "people involved in a traffic accident" such as *cyclist*, *driver* and *motorcyclist* (V + n), and also nouns for "body parts" such as *ankle*, *knee* and *wrist* (V + n). Including such nouns undoubtedly contribute to clarifying the verb's usage. Also, it makes it possible to include less frequent nouns that are not treated as

headwords such as “cyclist” and “motorcyclist.” OCD2 does not include them as headwords, so the collocations of “**injure** + cyclist” and “**injure** + motorcyclist” have no chance to appear in this dictionary.

Another advantage of MCD is that it consistently includes “and/or” category[7]. Let us look at some examples from MCD:

age (n) (sense 1)
and/or: *ability*, *background*, *ethnicity*, *gender*, *nationality*, *occupation*, …
angry (adj)
and/or: *aggressive*, *annoyed*, *irritable*, *depressed*, *sad*, …

From the collocates for the entry of **age**, for instance, we can assume that the coordination pattern of **age** is often used in the context of job requirements in such phrases as *age and ability* and *age and background*. Also, information for the entry of **angry** enables us to describe somebody’s anger in a more precise way. Thus, the pattern of “and/or” can be of great use for encoding purposes.

The final point to be considered as to the collocational patterns is about “phrases.” OCD2 and LCD have this category and put miscellaneous phrases here. For example, OCD2 includes *stand accused of* in the phrases category of **accuse** (v). This category makes it possible to list expressions that do not fall into any standard collocational patterns but are worth including. MCD does not use this category and tries to make its patterns as clear as possible. However, MCD’s policy may also cause a lack of information. For example, phrases such as *sb’s* ***nerves*** *are on edge* (=they feel slightly nervous or worried) (LCD) and *in your mind’s* ***eye*** (=your imagination) (OCD2) are certainly useful for learners, but they are not included in MCD due mainly to the lack of this category.

3.3.2. Groups of collocates

Collocates are grouped according to their meanings in each dictionary. In OCD2 and MCD, words are listed in alphabetical order within each group. OCD2 states that “[t]he groups are arranged in an order that tries to be as intuitive as possible” (viii). For example, adjective collocates for the entry of **behaviour** (n) are grouped as “*exemplary*, *good* | *acceptable* | *normal* |

bizarre, *strange*, *suspicious* | ..." ("|" indicates group boundaries) and the order of groups seems to be intuitive as the list starts from positive collocates and then followed by neutral and negative words; notice also that words are listed alphabetically within each group. However, each semantic group is not given any explanations in OCD2[8]. By contrast, MCD gives a brief meaning for each group, though the policy for the order of the groups is left unexplained. An example below is taken from the entry of **child** (n), where, again, collocates appear in alphabetical order within each set.

adj + N
of a particular age: *adolescent*, *little*, *newborn*, *preschool*, *school-aged*, *small*, *teenage*, *young*
with problems: *abused*, *at-risk*, *disabled*, *disadvantaged*, *neglected*, *special-needs*, *underprivileged*, *vulnerable*
badly behaved: *badly behaved*, *delinquent*, *difficult*, *naughty*, *problem*, *spoiled*

This strategy seems to be more user-friendly than OCD2's policy. However, as Coffey (2010: 339) points out, meaning descriptions do not always successfully function as sign posts. The above example includes such a case; the definition "badly behaved" is just a repetition of the first collocate.

LCD states in its preface that "[c]ollocations are listed in order of frequency, so that you can see the most common collocations first" (iv). Compare the following examples taken from the "v + N" sections in the entries of **cigarette** in each dictionary:

OCD2
smoke | *draw on*, *pull on*, *suck on* | *light* | *extinguish*, *put out*, *stub out* | *roll* | *flick*, *flick away* | *bum* | *advertise*
MCD
light or smoke a cigarette: *draw on*, *light*, *puff on*, *smoke*
finish a cigarette: *extinguish*, *finish*, *put out*, *stub out*
LCD
smoke a cigarette
light a cigarette
put out a cigarette also *extinguish a cigarette* formal (=stop it

burning)

pull on/drag on/draw on a cigarette (=smoke a cigarette with deep breaths)

roll a cigarette (=make your own cigarette using special paper)

LCD's entry tells us that the most frequent collocation is "smoke + **cigarette**," and also "pull on + **cigarette**" is the most common collocation in the meaning of 'smoking with deep breaths.' Rather than simply giving words in alphabetical order, this is far more informative for users. One possible problem of this policy is that similar words may appear separately. For example, at the entry of **citizen** (n), *a good citizen*, *a respectable/decent citizen* and *a model citizen* are listed separately[9], making it difficult to recognize the semantic closeness of these expressions.

Another advantage of LCD is that "[t]he meanings of difficult collocations are explained in brackets" (iv). This means that semantic annotation is not given to the group of collocates but to the specific phrases as can be seen above (the paraphrase starts with "="). This policy also makes it possible to annotate the meanings of idiomatic or semantically opaque phrases, or of genre-specific phrases such as *peer **review***, which is annotated as "(=in scientific and other studies, the examination of someone's work by other scientists, researchers etc)." Since the meaning of the phrase is not always the sum of each word, this surely helps users understand the meaning and the usage of the phrase.

3.4. Examples

Just like other learner dictionaries, the collocations dictionaries have examples in each entry. MCD and LCD give sentence examples whereas OCD2 has both sentence and phrase examples. Some random examples are given below:

clue (n) [prep] ~ *(as) to*: a ~ *as to* her whereabouts (OCD2)

departure (n) [v + N] *delay*: She'd heard that I'd had to *delay* my departure. (MCD)

citizen (n) [adj] *a good citizen*: The education system is designed to produce *good* citizens. (LCD)

LCD has the richest examples; almost every collocation is accompanied by an example. By contrast, MCD provides only one example for most semantic groups and OCD2 offers examples to only a small part of semantic groups; both dictionaries leave many of their collocations unexplained and are less attractive in this respect. Obviously, having more examples is an advantage especially for encoding purposes (see 5.2.2 and 5.2.3).

3.5. Labels

OCD2 and LCD employ various labels. The following are the labels listed on each dictionary's inside front cover and some random examples:

OCD2

[related to attitudes/situations] disapproving, figurative, formal, historical, humorous, informal, ironic, literary, offensive, old-fashioned, saying, slang, technical

[related to subject areas] biology, business, computing, economics, finance, law, etc.

[related to regions and institutions in particular countries] (esp.) AmE, (esp.) BrE, in Australia, in England and Wales, etc.

call (v) (sense 3) [adv] *free* (BrE), *toll-free* (AmE)

self (n) [adj] *good* (humorous, esp. BrE)

LCD

Ac (used to show that a word is on the Academic Word List), AmE, BrE, formal, informal, spoken, written, literary, technical, old-fashioned, disapproving, humorous

accident (n) [v] *be involved in an accident* formal

eye (n) [v + N] *avert your eyes* literary

OCD2 notes that "[l]abels used in the dictionary generally apply to the *collocation*—that is to the two words in combination—and not to the individual words" (inside front cover). To take "do drugs" as an example, "neither *do* nor *drugs* is informal in itself" (ix) but as a phrase it is assigned an "informal" label. The same rule is true in LCD, which states, "Labels tell you if the collocation is only used in formal or informal English" for the use of "formal" label (iv).

MCD does not mention its policy on its use of labels in the preface or

elsewhere, but it uses "informal" labels as can be seen in the following example:

recipe (n) (sense 1) [adj + N] *yummy* informal

However, compared to OCD2 and LCD, MCD is insufficient in this respect. Specifically, no attention is paid to regional varieties, which results in a serious defect of the dictionary. For example, according to OCD2 and LCD, "have a **guess**" is a Briticism whereas "take a **guess**" is an Americanism, but this kind of information is not found in the entry of **guess** in MCD which simply gives *have*, *hazard*, *make* and *take* as "v + N" collocates.

3.6. Summary

The microstructure of each dictionary is, in general, similar to each other, but there are some differences as Table 3.1 summarizes. It can be said that LCD is advantageous in many respects and, as far as the microstructure is concerned, this is one of the most powerful tools for users to produce natural English. (Section 5 will show that most of our survey participants actually preferred LCD.)

Table 3.1
Summary of Differences in the Microstructure

		OCD2	MCD	LCD
Definitions of Headwords		only for polysemous words	✓	✓
Collocational Patterns	Noun Collocates at Verb and Adjective Entries		✓	✓
	"phrases"	✓		✓
	"and/or"		✓	
	Frequency Order			✓
Order of Collocates		alphabetical	alphabetical	frequency

Meaning Descriptions of Collocations		occasional	for each semantic group	for semantically opaque phrases
Examples		for part of semantic groups	for each semantic group	for each phrase
Labels	Register	✓	"informal" only	✓
	Geographical	✓		✓

Since collocations dictionaries are based on corpora as stated in the preface in each of our target dictionaries (see also 4.2), it is not difficult to add frequency information (as in LCD), and register and genre information (as in OCD2 and LCD). Those dictionaries that do not take advantage of these kinds of information can be improved by using their corpora more extensively in future editions. Also, it might be possible to add more notes for learners using learner corpora[10]. The recent trends toward creating huge corpora will surely contribute to improving collocations dictionaries in the not-so-distant future.

(Section 3 by Uchida)

4. Comparison of collocates given in three collocations dictionaries

This section compares the collocates given under several same headwords in each dictionary especially in terms of their frequency. We will show the quantitative summaries of the collocates of five sampled headwords (4.1), and give an overview of those collocates' frequency in large corpora to verify whether highly frequent collocates are thoroughly included in our target dictionaries (4.2).

The five sample headwords considered in this section were selected taking the following steps:

1. We picked the first headwords, excluding empty ones only for cross reference, starting with each alphabet on the print edition of LCD whose coverage is the smallest among the three dictionaries: **abandon** (v), **baby** (n), **cab** (n), **dam** (n), **ear** (n), and so on.

2. Only those headwords covered in all three dictionaries were kept. This step excluded seven items, and the 17 remaining words include **abandon** (v), **cab** (n), **dam** (n), **face** (n), **gadget** (n), etc.
3. Following the rough proportion of the word classes in the three dictionaries—20% for adjectives, 60% for nouns and 20% for verbs (see Table 2.1)—we decided to pick one adjective, three nouns and one verb. We had only one adjective (**ugly**) and one verb (**abandon**) after the second step above, and they were automatically chosen as part of our samples.
4. To select three nouns out of 15, we decided to choose those headwords which have only one sense in the paper version of LCD for ease of comparison, which narrowed the number to 11. Next, we decided to use the three most frequent items in the BNC. However, the most frequent among the 11 words, **face**, is divided into two senses in OCD2 where collocates of this word are given in separate sections. They are given collectively under one sense in MCD and LCD, so we avoided this word, and finally chose the nouns **habit**, **qualification** and **wage**.

4.1. Comparison of the number of collocates across the three dictionaries

To compare the coverage in different dictionaries, we first collected all the collocates given under the headwords **abandon** (v), **habit** (n), **qualification** (n), **ugly** (adj) and **wage** (n) in all three dictionaries. We checked both print and electronic versions for LCD and OCD2 and included those items that are only in the electronic versions. For **abandon**, we only looked at the sense of 'stop doing/supporting something' because this is the only sense covered in MCD. Likewise, we only checked the 'credentials' sense for **qualification**. For **ugly**, we focused on the 'unattractive' sense given as sense 1 in LCD for brevity's sake. The summaries of the number of collocates are given in Tables 4.1-4.5.

Table 4.1
Number of Collocates Given for **abandon** (v)

	OCD2 (sense 2)	MCD	LCD[1)] (sense 2)
abandon + n		8	10
adv	18	12	9
v + **abandon**	3		2

phrases			1
abandon + prep	2		
(Total)	23	20	22

Table 4.2

Number of Collocates Given for **habit** (n)

	OCD2	MCD	LCD
adj/n + **habit**	61	21	20
v + **habit**	19	13	10
habit + v	1		
prep + **habit**	3		2
phrases	5	4[2)]	3
(Total)	89	38	35

Table 4.3

Number of Collocates Given for **qualification** (n)

	OCD2 (sense 1)	MCD	LCD
adj/n + **qualification**	21	41	15
qualification + n	1		
v + **qualification**	8	21	6
prep	2		4
and/or		7	
(Total)	32	69	25

Table 4.4

Number of Collocates Given for **ugly** (n)

	OCD2 (sense 1)	MCD (sense 1)	LCD (sense 1)
ugly + n			16
adv + **ugly**	8	10	5
v + **ugly**	3		1
phrases			2
(Total)	11	10	24

Table 4.5

Number of Collocates Given for **wage** (n)

	OCD2	MCD	LCD
adj + **wage**	31	11	10
v + **wage**	23	15	11
wage + v	3		3
wage + n	32	8	10
phrases	8		6
(Total)	97	34	40

From the quantitative summaries in Tables 4.1-4.5 alone, no particular dictionary can be said to offer more collocates consistently. More extensive research needs to be carried out in order to see how wide a range of collocates each dictionary offers.

4.2. Frequencies of collocates

We have checked the frequencies of each collocate, except for multi-word items, in the BNC and the Corpus of Contemporary American English (COCA). Part of the result is laid out in Tables 4.6-4.7. In these tables "✔" signifies that the collocate is given in the dictionary in question. The "BNC/COCA Rank" shows the rank of the collocate among the possible collocates in the specified syntactic pattern; for example, "attempt(s)" is the third most frequent item when we search for the nouns within the span of 1-3 words to the right of "abandon" and its inflectional forms in the BNC. The sign "—" means the collocate is used less than five times in the corpus.

Table 4.6

Coverage and Frequencies of Collocates for **abandon** + n

	OCD2	MCD	LCD	BNC Rank	COCA Rank
attempt		✔	✔	3	22
belief		✔		25	69
decision			✔	—	—
effort			✔	16	5
faith		✔		41	40
hope		✔	✔	5	9
idea		✔	✔	1	2
plan		✔	✔	2	3
policy			✔	6	13

project			✔	12	10
pretence		✔		15	56
principle		✔	✔	11	16
search			✔	28	44

Table 4.7

Coverage and Frequencies of Collocates for v + **habit**

	OCD2	MCD	LCD	BNC Rank	COCA Rank
acquire	✔	✔		10	16
adopt	✔	✔		16	18
alert		✔		—	—
become	✔		✔	6	7
break	✔	✔	✔	7	10
change	✔	✔	✔	5	3
cultivate	✔	✔		—	47
develop	✔	✔	✔	8	6
establish	✔			12	20
form	✔		✔	11	19
have	✔	✔	✔	1	2
kick	✔	✔	✔	9	8
make	✔		✔	4	4
overcome		✔		—	109
quit		✔		—	58
support	✔			—	10

The results of our survey shows that the three collocations dictionaries extensively cover frequent collocates, with only a few exceptions such as "claim(s)" for "**abandon** + n," which is the seventh most frequent item in the BNC. This result may seem quite reasonable because all of these dictionaries are corpus-based and will conform to our expectation that they offer the users frequent collocates of a given word, but we have to remember that there is no dictionary that covers all collocates found in other dictionaries. Also, while it is true that their policy to offer highly frequent collocates for frequent headwords makes sense, the users of collocations dictionaries are likely to use them for encoding purposes more often than for decoding. Considering this fact, users are likely to face problems when they want to

know the collocates of infrequent words which will be covered only on a small scale, if any, or when they want to freely express a variety of ideas that may not necessarily be covered within the frequent collocates extracted from corpora. In other words, all our target dictionaries focus on very frequent headwords and collocates, but this could also be a limiting factor for learners.

(Section 4 by Ishii and Kobayashi)

5. Users' perspective

In this section we will highlight the users of collocations dictionaries. Most of us will agree on the effectiveness of collocations dictionaries for learners, but it does not necessarily mean that every learner can use these dictionaries effectively. We conducted a questionnaire survey to make clear how properly learners can use collocations dictionaries and how learners find this special kind of reference work.

5.1. Participants and questions

The participants of our survey are 64 lower intermediate non-English majors at a Japanese university. They are familiar with the notion of collocations in their English classes and take a quiz on collocations in every class.

This survey consists of three sections: A, B and C. In section A, the participants were presented color photocopies of relevant entries in three dictionaries indicated as Dictionary X/Y/Z which corresponds to LCD/MCD/OCD2 respectively, and were asked to write an appropriate expression which is equivalent to a Japanese phrase by referring to the entries. The headwords whose entries our participants had to check to write their answer were carefully chosen; we made sure that the headword and the collocate which was supposed to be an answer were in all three dictionaries. We asked our participants to carefully read all the relevant entries from three dictionaries even if they knew the answer without checking them. Then, the participants were required to decide which dictionary helped most when they answered the question. Also, they were requested to choose or freely write the reason(s) why they thought the dictionary they chose was most helpful. The choices for Questions A1-3/A2-3/A3-3 are based on the result of a previous study (Kawamura and Ishii 2013: 51-52). The original questions and choices are written in Japanese,

but their English translations are as follows:

A1-1: Refer to the entries of **exam** in each dictionary and write a correct expression in English equivalent to a Japanese expression meaning 'I need to retake this exam.': I need to _____.

A1-2: Which dictionary was most helpful to answer the question above? (multiple choices allowed)

a. Dictionary X
b. Dictionary Y
c. Dictionary Z
d. No dictionary was helpful, and I cannot find the right answer.

A1-3: Choose the reason why you found that dictionary helpful. (multiple choices allowed)

a. The examples are useful.
b. The amount of information is appropriate.
c. A meaning is given for each phrase or set of collocates, such as "(=look at different websites)" for *surf the* ***internet*** in Dictionary X and "have or use the Internet" for the set of collocates *access*, *be connected to*, *be linked to*, *connect to*, *go on*, *link to* and *use* for **the Internet** in Dictionary Y.
d. It is easy to read in terms of the use of boldface, italics, line breaks, colors, and so on.
e. Others: Write the reason.

A2-1: Refer to the entries of **Internet/internet** in each dictionary and write a correct expression in English equivalent to a Japanese expression meaning 'I ordered the book over the Internet.': I ordered the book _____.

A2-2: (the same as A1-2)

A2-3: (the same as A1-3)

A3-1: Refer to the entries of **health** in each dictionary and write a correct expression in English equivalent to a Japanese expression meaning 'Stress affects our mental health.': Stress affects our _____.

A3-2: (the same as A1-2)

A3-3: (the same as A1-3)

In section B of our survey, we asked the participants to choose the

features of the dictionary in question which they find superior/inferior to those of the other dictionaries. The choices are based on the results of a previous study (Kawamura and Ishii 2013: 52-55) and common for all the following questions:

B1: Choose the features of Dictionary X you find superior to those of the other two dictionaries. (multiple choices allowed)

B2: Choose the features of Dictionary X you find inferior to those of the other two dictionaries. (multiple choices allowed)

B3: Choose the features of Dictionary Y you find superior to those of the other two dictionaries. (multiple choices allowed)

B4: Choose the features of Dictionary Y you find inferior to those of the other two dictionaries. (multiple choices allowed)

B5: Choose the features of Dictionary Z you find superior to those of the other two dictionaries. (multiple choices allowed)

B6: Choose the features of Dictionary Z you find inferior to those of the other two dictionaries. (multiple choices allowed)

a. The number of collocates is (not) appropriate.
b. A meaning is (not) given for each phrase or set of collocates. (present in Dictionaries X and Y)
c. The number of examples is (not) appropriate.
d. It is (not) easy to read in terms of the use of boldface, italics, line breaks, colors, and so on.
e. Others: Write the feature.
f. Nothing in particular.

Finally, in section C, the participants were asked the following question:

C: Which improvements would make collocations dictionaries easier to use? (multiple choices allowed)

a. more collocates
b. more examples
c. English explanations for proper choice of collocates within each set of collocates
d. frequency of each collocate

e. Japanese translations or explanations

f. Others: Write your suggestion.

5.2. Results

5.2.1. Results for section A

The percentages of correct answers to each question are given in Table 5.1. The result for Question A1-1 is very poor (10.9%) mainly because many participants chose a wrong determiner and wrote "retake an exam" while the Japanese expression reads "retake this exam." The plausible reason for this error is that they found "*retake an exam* (also *resit an exam* BrE)" in LCD and just copied it. If we allow this minor error and focus on the correct choice of a verb collocate, then 60.9% of the participants gave a correct answer, which is indicated with an asterisk in Table 5.1. The results show that more than half of the participants answered the questions correctly in Questions A2-1 and A3-1, and chose a correct verb in Question A1-1. Also, almost all participants wrote a correct part-of-speech word even if they were wrong; we can assume that they at least understood which part of the dictionaries they had to look at. However, considering the fact that the answers are present in all three dictionaries, these percentages should be higher. We need to recognize that not a small number of our participants cannot properly use these collocations dictionaries for some reasons, part of which will be considered below.

Table 5.1

Percentages of Correct Answers for Each Question

Question	Answers Treated as Correct	Percentage of Correct Answers
A1-1	retake/resit this/the exam	10.9% / 60.9%*
A2-1	on/over/via the Internet/internet	87.5%
A3-1	mental/emotional/psychological/spiritual health	64.1%

The results for questions asking the participants which dictionary was most helpful to answer the questions A1-1/A2-1/A3-1 are summarized in Table 5.2. Note that only those who were correct in the previous question were counted, and those who chose a correct verb but used a determiner that

didn't match the Japanese phrase given are counted separately and shown with an asterisk. (The same policy applies to Table 5.3.) As these results clearly show, most of the participants chose LCD as the most helpful among the three dictionaries.

Table 5.2
Dictionary Found Most Helpful (Multiple Choices Allowed)

Question	Z [OCD2]	Y [MCD]	X [LCD]	None
A1-2 (retake this exam)	0 (0%)	1 (14.3%)	6 (85.7%)	0 (0%)
A1-2 (retake an exam)*	1 (3.1%)	1 (3.1%)	31 (96.9%)	0 (0%)
A2-2 (over the Internet)	3 (5.4%)	2 (3.6%)	53 (94.6%)	2 (3.6%)
A3-2 (mental health)	1 (2.4%)	8 (19.5%)	34 (82.9%)	2 (4.9%)

(The percentages are of those who gave a correct answer.)

Table 5.3 summarizes the features which the participants chose as the reason(s) why they thought the dictionary of their choice was the most helpful. The results indicate that many of our participants felt helped by the examples and easiness to read the dictionary. On the contrary, our participants did not indicate that the meanings for each set of collocates were such an important factor when they found the answers, which is a little surprising because we expected this feature found in MCD and LCD to be highly helpful to users.

Table 5.3
Reasons for Finding the Dictionary Helpful (Multiple Choices Allowed)

Question	Examples	Amount of Information	Meanings	Easy to Read	Others
A1-3 (retake this exam)	6 (85.7%)	1 (14.3%)	1 (14.3%)	6 (85.7%)	0 (0%)
A1-3 (retake an exam)*	13 (40.6%)	6 (18.8%)	7 (21.9%)	20 (62.5%)	2 (6.3%)
A2-3 (over the Internet)	31 (55.4%)	14 (25.0%)	9 (16.1%)	40 (71.4%)	4 (7.1%)
A3-3 (mental health)	21 (51.2%)	15 (36.6%)	9 (22.0%)	25 (61.0%)	2 (4.9%)

(The percentages are of those who gave a correct answer.)

5.2.2. Results for section B

Table 5.4 shows which features of each dictionary were judged to be superior/inferior to the others by our participants. In general, LCD is deemed to be more useful than MCD and OCD2 in many ways, but what is the most striking difference across these dictionaries is that LCD was praised for being much easier to read than the others. We can easily notice that there are at least three advantages in how the information is presented in LCD. First, there is a line break after eachitem, which yields more white space. Although this also holds true of MCD, the sets of collocates in MCD are larger than the items in LCD in most cases, which makes MCD more crowded with words and with less white space. Second, LCD uses larger font sizes than the other two dictionaries. And third, the syntactic pattern such as "ADJECTIVES" and "VERBS" stand out well in entries. These characteristics give us the impression that this dictionary is much less packed and make it easier to look for the information we want to obtain than in MCD or OCD2. However, it must be kept in mind that in this survey the photocopies of print editions were used, and if the CD-ROM version of OCD2 had been offered, the criticism would have been less severe.

Table 5.4
Strengths and Weaknesses in Each Dictionary (Multiple Choices Allowed)

	Num. of Collocates	Meanings	Num. of Examples	Easy to Read	Others	Nothing
Strengths of LCD	37	27	33	61	10	0
Weaknesses of LCD	20	6	7	1	3	37
Strengths of MCD	33	15	18	3	1	21
Weaknesses of MCD	20	11	7	52	7	5
Strengths of OCD2	33	9	12	6	1	25
Weaknesses	19	16	19	52	6	3

of OCD2						

(The total number of participants is 64.)

As for the number of collocates, there seems to be no big difference in the users' evaluation of the three dictionaries. This may be, however, due to the nature of our survey; appropriate collocates are included in all the dictionaries. Regarding the meanings of each set of collocates, about a quarter to a third of the participants answered that they are useful, although this feature was not regarded to be so important when they tried to find a correct collocate (see Table 5.3). Concerning the number of examples, it seems many users feel that they will benefit if more examples are offered (see also Table 5.5 below).

5.2.3. Results for section C

Table 5.5 presents the result for the question: "Which improvements would make collocations dictionaries easier to use?" As for the numbers of collocates in our target dictionaries, our participants seem satisfied with them, although this could be due to the fact that they didn't experience the lack of the items they were looking for in our survey. On the other hand, around a third to half of the participants answered the other four improvements—more examples, English explanations for proper choice of collocates within each set of collocates, frequency of each collocate and Japanese translations or explanations—would be beneficial, although the last would be impossible in monolingual dictionaries. All of these four possible improvements are related to choosing the best collocate, and they must be related to the not-so-high percentages of correct answers for questions A1-1/A2-1/A3-1. Most of the participants who chose "Others" wrote that the easiness to read is the most important factor.

Table 5.5

Features Which Our Participants Felt Would Make Collocations Dictionary Easier to Use (Multiple Choices Allowed)

More Collocates	More Examples	English Explanation for Proper Choice of Collocates	Frequency	Japanese Translations /Explanations	Others
9	28	29	25	26	16

(The total number of participants is 64.)

5.3. Summary

Through our survey, we have revealed three important aspects, among others, in collocations dictionaries from the viewpoint of users. First, the easiness to read weighs importantly with the users. It is hard especially for lower intermediate learners to find an appropriate collocate and check examples written in a foreign language on a cramped page. Second, many users rely on examples to find appropriate collocates. Third, the meanings for each set of collocates in LCD and MCD are helpful to not a small proportion of users. These can be seen as keys for improvement in potential future editions or for their potential competitors.

(Section 5 by Hayashi and Ishii)

6. Conclusion

This paper briefly reviewed and compared three collocations dictionaries for intermediate to advanced learners of English. In Section 2, we compared their headwords giving the actual numbers and showed the advantage of OCD2 in terms of the number of headwords while MCD and LCD have their own characteristic features. Section 3 dealt with the microstructures and compared their different policies especially on the arrangement of collocates, on examples and on labels. Section 4 focused on the collocates and confirmed that in general all three dictionaries offer highly frequent collocates based on corpora. In Section 5, we described how these dictionaries are used by lower intermediate learners and what aspects they find useful or needing improvement. Through our review, it has become clear that every collocations dictionary has its own advantages compared to other competitors, but at the same time they leave room for improvement to be more user-friendly to a wider range of learners.

Notes

Section 1

1) LCD does not differentiate between “collocations” and “collocates” and collectively calls them “collocations.”

2) There are two entries of **peep** (v) in the electronic version, one of which has an additional example. The figure of 22,864 is the result of counting the

two entries as one item.

Section 2

1) We would like to express our gratitude to Seijo University for funding the research for this section through the Grant-in-Aid for Creative Research in 2013-2014.

2) Headwords are written in bold face in this paper.

3) Different dictionaries employ different policies on how to represent parts of speech. In this paper, we use the simplest type of notations such as "(n)" and "(v)" for all three dictionaries.

4) The number of empty headwords is 104, and even if we include them, the total number is 8,522.

5) The number for adjectives includes 11 "adj., adv.", 14 "adj., noun" and one "adj., pron.", and the number for nouns includes one "noun, exclamation."

6) 113 phrasal verbs are included in this figure.

7) The number for adjectives includes 11 "adj, adv" and 19 other similar items that are given more than one part of speech. The same process of lumping together has been carried out for nouns and adverbs as well.

8) This number includes 34 other items including five items labeled as "conjunction" and several other types.

9) For those items that have more than one part of speech, we used the first part of speech for our counting; for example, an item classified as "adj., noun" was dealt with as an adjective.

10) Phrasal verbs in MCD are counted as verbs.

11) We used BNC*web*'s "Frequency lists" function.

12) OCD2 has 11 headwords labeled as "adj., adv."

Section 3

1) The IPA transcriptions and audio clips for headwords are available on the online version of LCD. Although it is desirable to include phonetic information of the headwords (and also of the collocates), this policy can be justified considering the purpose of collocations dictionaries, i.e., serving learners who want to write naturally.

2) For the sake of simplicity, the same notation is used for all three dictionaries. Capitalized letters indicate the position of headwords. N/n, V/v, ADJ/adj and adv stand for noun, verb, adjective and adverb respectively.

3) MCD has some other patterns such as "V + with," "ADJ + in" and "n + of + N."
4) LCD also has some other minor patterns such as "COLOUR" and "SHAPE/POSITION" for **eye** (n).
5) LCD also includes predicative adjectives in this category. For example, at the entry of **prediction** (n), the following expressions can be found: *an accurate prediction*, *a prediction is correct/right*, *a prediction is wrong/incorrect* and *a reliable prediction*.
6) The preface states as follows: "The third question asked (Would a student look up this entry to find this expression?) led to the exclusion of noun collocates from verb and adjective entries" (vi).
7) OCD2 and LCD's "phrases" category sometimes gives phrases of this pattern; *angry and frustrated/upset* is given in the entry of **angry** in LCD.
8) For some collocations and idiomatic phrases, OCD2 gives a simple paraphrase: "*the* ***burden*** *of proof* (=the responsibility of proving that sth is true)," for instance.
9) The latter two collocations are included only in the electronic version.
10) LCD provides such notes in some entries. For example, at the entry of **research** (n) we can find a note "Don't say 'make research'."

Section 4
1) The number of collocates, not items, are counted; for example, the item ***abandon*** *an attempt/effort* is treated as giving two noun collocates. The only exception is the "phrases" category where each item is counted as one phrase; *change/break the* ***habits*** *of a lifetime* is counted as one item, for instance.
2) These four items are given as phrases in the note at this entry; MCD states that it uses notes "when there is a common way of expressing the same idea using a phrase rather than a collocation" (xii).

Dictionaries Referenced

LCD: *Longman Collocations Dictionary and Thesaurus*. Harlow: Pearson Education, 2013.
MCD: *Macmillan Collocations Dictionary*. Oxford: Macmillan Education, 2010.
OCD2: *Oxford Collocations Dictionary*. 2nd ed. Oxford: Oxford University

Press, 2009.

OALD8: *Oxford Advanced Learner's Dictionary of Current English*. 8th ed. Oxford: Oxford University Press, 2010.

Other Works Cited

Coffey, Stephen. "A New Pedagogical Dictionary of English Collocations." *Internal Journal of Lexicography* 24 (2010): 328-341.

Kawamura, Akihiko, and Yasutake Ishii. "Komyunikeeshon Nouryoku Ikusei no Tame no Goi Shidou—Poraitonesu to Korokeeshon no Kanten kara—[Developing Students' Vocabulary Knowledge for Better Communication Skills: With Special Emphasis on Politeness and Collocation]." *Social Innovation Studies* 8.2 (2013): 37-68.

Macmillan Education. *Macmillan Education English Language Teaching Catalogue 2013*. 2013.

Oxford University Press. *Oxford English Language Teaching Catalogue 2013*. 2013.

Pearson ELT. *The Pearson ELT and Penguin Readers Catalogue 2013*. 2013.