
A Critique of the Controlled Defining Vocabulary in *Longman Dictionary of Contemporary English*

Hai Xu, *Centre for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China (xuhai1101@yahoo.com.cn)*

Abstract: This article critically analyses the Longman Defining Vocabulary (LDV) in relation to its size, range and frequency, senses, parts of speech, affixes, and multiword expressions. The recent versions of the LDV contain a relatively fixed number of items. Over 85% of those items were found to be highly frequent, and for the defining purpose, genus terms, grammatical terms, etc. have also been included. The number of affixes in the LDV has been greatly reduced, and some common derivatives have been listed separately. On the other hand, the actual size of the LDV is much larger than was reported, for LDOCE did not distinguish between the LDV items with different senses or forms. It was found that the claim of using the 'most common meanings' of the LDV items is not always held true. The parts of speech of the LDV items have not been systematically indicated. Many multiword expressions, which have been used in the definitions in LDOCE, are not part of the LDV. This study sheds some light on the improvement in the practice of using a controlled defining vocabulary in an English learner's dictionary.

Keywords: CONTROLLED DEFINING VOCABULARY, LONGMAN DEFINING VOCABULARY, SIZE, RANGE, FREQUENCY, SENSE INDICATION, POS INDICATION, INCLUSION OF AFFIXES, INCLUSION OF MULTIWORD EXPRESSIONS

Opsomming: 'n Beoordeling van die beperkte definisiewoordeskate in die *Longman Dictionary of Contemporary English*. In hierdie artikel word 'n kritiese ontleding gedoen van die Longman Definisiewoordeskate (LDW) met betrekking tot die grootte, omvang en frekwensie, betekenis, woordsoorte, affikse en meerwoordige uitdrukkings daarvan. Die onlangse weergawe van die LDW bevat 'n relatief vaste aantal items. Daar is bevind dat meer as 85% van hierdie items 'n hoë frekwensie het, en dat genusterme, grammatiese terme, ens., ook ingesluit is vir die doeleindes van definiëring. Die aantal affikse in die LDW is grootliks verminder, en sommige algemene afleidings is apart gelys. Aan die ander kant was die werklike grootte van die LDW baie groter as wat aangedui is, aangesien LDOCE nie onderskei het tussen die LDW-items met verskillende betekenis of vorms nie. Daar is bevind dat die bewering dat die 'algemeenste betekenis' van die LDW-items gebruik is, nie altyd gegeld het nie. Die woordsoorte van die LDW-items is nie sistematies aangedui nie. Baie meerwoordige uitdrukkings wat in die definisies van LDOCE gebruik is, vorm nie deel van die LDW nie. Hierdie studie werp lig op die verbetering in die gebruik van 'n beperkte definisiewoordeskate in 'n Engelse aanleerderswoordeboek.

Sleutelwoorde: BEPERKTE DEFINISIEWOORDESKATE, LONGMAN DEFINISIEWOORDESKATE, GROOTTE, OMVANG, FREKWENSIE, BETEKENISAANDUIDING, WOORDSOORTE

1. Introduction

In the English learner's dictionary market, the five major competitors except for COBUILD have all acknowledged the role of a controlled defining vocabulary (CDV) in the decoding task, and compiled definitions on the basis of around 2,000 to 3,500 CDV items. The practice of using a CDV in definitions originates from the *New Method English Dictionary* (NMED), a learner's dictionary which had 'an entirely original feature — definitions based on a "minimum adequate vocabulary"' (Cowie 1999: 24). As a response to the vocabulary control movement during the late 1920s and early to mid-1930s, NMED used 1,490 words to define 23,898 vocabulary items. The practice of using a CDV had not been continued until the birth of LDOCE in 1978. LDOCE used a CDV of approximately 2,000 items. While for the first four editions of OALD, no attempt was made to set up a restricted defining vocabulary (Cowie 1999), the editorial policy of OALD5 changed radically. OALD5 used some 3,500 CDV items. The number of CDV items in OALD6 was claimed to have fallen to under 3,000 words, and they were called the 'Oxford 3000' in OALD7 and OALD8. CIDE and its successive editions — CALD — restricted the use of a defining vocabulary to less than 2,000 words. MED, a relatively new addition to the English learner's dictionary family, used a CDV of under 2,500 words to write definitions.

The strength of a CDV lies in its comprehensibility and simplicity. The study of MacFarquhar and Richards (1983) shows that over half of their subjects judged the definitions in LDOCE1 more helpful and easier to understand than those in either OALD3 or WNWD2. Herbst's study (1986) also demonstrates that LDOCE1 was far more successful in providing easily comprehensible and sufficiently accurate definitions than OALD3 and CULD. A CDV is believed to lighten the learning burden of foreign students and to facilitate their decoding task. Even for native speakers, some definitions in a learner's dictionary are considered to be more comprehensible than those in a collegiate dictionary. McCreary and Amacker (2006) reported that in a comprehension task of hard words, the groups of American college students who used an advanced learner's dictionary (i.e. MED1) performed slightly better than the groups using a collegiate desk dictionary (i.e. MW11). One of the reasons that some MW11 entries led students to induced errors lies in that they used low frequency words in the defining language.

On the other hand, there are some criticisms of CDV. Some definitions using CDV items are deemed to be oversimplified, vague, unnatural or convoluted (Stein 1979: 6; MacFarquhar and Richards 1983: 115; Béjoint 2000: 70; Fontenelle 2009: 419-420). Some CDV items may achieve simplicity at the expense of accuracy. The lack of preciseness is particularly conspicuous for

names of animals, plants, substances, and games.' (Stein 1979: 6) In order to avoid non-'core' vocabulary, the lexicographers of LDOCE1 sometimes had to resort to 'syntactically more complex, convoluted (or less natural) constructions' (Fontenelle 2009: 419). For instance, instead of defining **tabasco** as 'a very hot sauce [...]', LDOCE1 compiled an unnatural definition like 'a very hot-tasting liquid [...]', for **sauce** was not part of the controlled vocabulary.

In the literature, the CDVs, particularly the Longman Defining Vocabulary (LDV), have been critically examined in relation to their size, range and frequency, senses, parts of speech, affixes, multiword expressions, etc. (Stein 1979; Michels and Noël 1984; Neubauer 1984; Herbst 1986, 1996; Jansen, Mergeai and Vanadroye 1987; Whitcut 1988; Fox 1989; Bogaards 1996, 2003, 2008; Cowie 1999; Clark 2003; Bullon and Leech 2007; Fontenelle 2009). As a follow-up, this study will focus on those aspects in the recent LDV, and consider its developments over its past five versions and in comparison with other CDVs.

2. Size of the LDV

Each version of the LDV included approximately 2,000 vocabulary items: the LDV1 contained 2,215 word forms¹; LDV2, 2,244; LDV3, 2,091; LDV4, 2,109; and LDV5, 2,107. Compared with the LDV1, the LDV2 removed 203 items, and added 232 new items. On the basis of the second version, the third deleted 359 items, and entered 206 new items. There were slight changes between the third and fourth versions: 25 items were excluded, and 43 were included. The changes between the fourth and fifth were the smallest: only one word (i.e. *thousandth*) and one phrasal verb (i.e. *look sth up*) were removed.

Of the 230 defining vocabulary items that occurred only once in the five versions of the LDV, 169 items (73.5%) were found in the first version, 60 (26.1%) in the second, and 1 (0.4%) in the third. The changes of defining vocabulary items over the last three versions were negligible.

The above results indicate that the core of the LDV items has been established. Actually, there are 1,769 defining vocabulary items (including 30 affixes) in common in the five versions of LDV.

However, it should be pointed out that the figure of around 2,000 may not have such a magic power. The actual size of a CDV is often underestimated. 'Partly in order to suggest that the learning burden represented by a CV is lighter than in fact it is, there is a tendency on the part of their designers to conceal their actual size.' (Cowie 1999: 110) The count of CDV items in a learner's dictionary is often fraught with some problems: not distinguishing between words with different senses or forms, not indicating parts of speech, and not including multiword expressions. That is why Stein (1979: 6) argued that, 'it is more important that this vocabulary be semantically self-sufficient than that it be restricted to an arbitrary fixed number of items'.

3. Range and frequency of the LDV items

To determine the range and frequency of the LDV items, the present study used the computer programs RANGE and FREQUENCY, designed by Heatley, Nation and Coxhead (2002), to compare the LDV lists with the BNC Word Family Lists. The programs 'can be used to find the coverage of a text by certain word lists' (Heatley, Nation and Coxhead 2002), and the BNC Word Family Lists reflect the nature of language use of typical English native speakers.

Table 1 shows the range and frequency of the LDV items² in comparison with the BNC Word Family Lists. In the five versions of the LDV, over 45% of the vocabulary items belong to the most frequent 1,000 words of English (i.e. the Base Word List 1 of the BNC Word Family Lists). More than 26% of the LDV items fall into the second range of the 1,000 most frequent words of English (i.e. the Base Word List 2). The LDV lists contain about 12% of the items which are in the range of the third 1,000 most frequent words of English (i.e. the Base Word List 3). To put it simply, over 85% of the items in the LDV lists are extremely frequent. Therefore, most of the items in the LDV lists should be familiar to dictionary users.

Table 1: Range and frequency of the LDV items

Word Lists	LDV1 (1978)	LDV2 (1987)	LDV3 (1995)	LDV4 (2003)	LDV5 (2009)
Basewrd 1	994 (45.3%)	1067 (47.8%)	1080 (52.0%)	1079 (51.5%)	1078 (51.5%)
Basewrd 2	583 (26.6%)	606 (27.2%)	605 (29.1%)	607 (29.0%)	607 (29.0%)
Basewrd 3	304 (13.9%)	296 (13.3%)	250 (12.0%)	255 (12.2%)	255 (12.2%)
Subtotal	85.8%	88.3%	93.1%	92.7%	92.7%
Not in the list	311 (14.2%)	262 (11.7%)	144 (6.9%)	155 (7.3%)	154 (7.3%)
Total	2192	2231	2079	2096	2094

We further conducted a study to find out which items in the LDV5 are outside the first three base word lists of the BNC. Running the *Mark texts* option in the RANGE program, we can easily identify those words. Altogether 154 items were found (See Appendix 1). Those items include 25 affixes (e.g. *-al, -ance, -ation, dis-, -ence*), 12 grammatical or metalanguage terms (e.g. *adjective, noun, particle, abbreviation, singular*), some genus terms (e.g. *cattle, creature, military, mineral, profession*), some words on science and technology (e.g. *atom, biology, network, software, spacecraft*), some emotional words (e.g. *anxiety, deceive, loyal, romantic, ugly*), and some common words on people (e.g. *girlfriend, opponent, poet, president, priest*), on animals (e.g. *beak, goat, lion, monkey, snake*) and on food (e.g. *corn, flour, onion, rice, sauce*), etc.

The inclusion of those words in the LDV demonstrates that high-frequency words alone are not adequate for the defining purpose (cf. Michels and Noël 1984; Neubauer 1984; Whitcut 1988). To explain word meanings, lexicographers need to use some genus or high-level words, high-generality words and a set of grammatical terms.

4. Senses of the LDV items

One of the main problems with the current CDV lists is that they do not indicate which senses a CDV item are included, and which are excluded. The LDV items have not been differentiated according to their senses. Although LDOCE claimed to use 'only the most common meanings', one is not always certain about the so-called 'most common meanings' of an LDV item.

Consider how the LDV item *note* was used in the definitions in LDOCE5. The item was used 180 times, which respectively correspond to 7 senses as indicated in LDOCE5. Table 2 lists the senses of *note* and the frequency of each usage.³

Table 2: Use of the LDV item *note* in the definitions

Meaning of the word <i>note</i>	Instance of usage in the definitions	Frequency of usage
1 TO REMIND YOU	<i>jottings</i> <i>n</i> [plural] <i>informal</i> short <u>notes</u> , usually written to remind yourself about something	14
2 FOR STUDYING	<i>review</i> ² <i>v</i> [T] 3 <i>AmE</i> to look again at something you have studied, such as <u>notes</u> , reports etc	5
3 SHORT LETTER	<i>leave</i> ¹ 8 [T] to deliver a message, <u>note</u> , package etc for someone or put it somewhere so that they will get it later	3
4 OFFICIAL LETTER	<i>chit</i> <i>n</i> [C] <i>BrE</i> 1 an official <u>note</u> that shows that you are allowed to have something	8
5 ADDITIONAL INFORMATION	<i>gloss</i> ² <i>v</i> [T] to provide a <u>note</u> in a piece of writing, explaining a difficult word, phrase, or idea	7
6 MUSIC	<i>pizzicato</i> <i>n</i> [U] musical <u>notes</u> played by pulling on the strings of an instrument	134
7 MONEY	<i>legal tender</i> <i>n</i> [U] coins or bank <u>notes</u> that people can officially use as money in a particular country	9

Table 2 indicates that the meaning of 'the sound/sign in music' predominates in the usage of *note* in the definitions, accounting for 74.4% of all the occurrences. In other words, the sixth sense of *note* rather than the first one has the highest frequency of usage in the definitions. Since an English learner's dictionary generally arranges word senses according to their frequency of usage, the sixth sense of *note* in LDOCE5 should not be regarded as the most common meaning.

The above phenomenon also exists in some items printed in small capitals⁴. Consider the use of the word *command* in the following definitions:

execute *v* 4 **COMPUTER** *technical* if a computer executes a program or **COMMAND** (= instruction), it makes the program or command happen or work

execution *n* 6 [C, U] *technical* when you make a computer program work, or a **COMMAND** (= instruction) happen

toolbox *n* [C] 2 a set of **COMMANDS** or **FUNCTIONS** which do various things in a computer program

user interface *n* [C] how a computer program looks on screen and how the user enters **COMMANDS** and information into the program

The above entries show all the usage of the small capital **COMMAND** in the definitions in LDOCE5.⁵ Clearly, **COMMAND** refers to 'an instruction to a computer to do something'. It corresponds to Sense 3 in LDOCE5. As for the first sense 'the control of a group of people or a situation', there is not an instance of usage. Hence, the claim of using the most common meanings of a defining vocabulary item in LDOCE is not always held true.

Some researchers thus suggested that CDV items should be cross-referenced to their appropriate senses in the dictionary (Neubauer 1984: 120; Herbst 1986: 114; Jansen, Mergeai and Vanadroye 1987: 84). In that case, however, the size of a CDV would rise substantially. A vocabulary item is more often than not polysemous, and only 20.19% of the LDV items are monosemous (Jansen, Mergeai and Vanadroye 1987: 84). As Petöfi (1977) pointed out, 'West's 1480 [*sic*, 1490] word forms of the defining vocabulary correspond to 4607 senses if one only counts the senses indicated by West himself in his dictionary' (cited from Neubauer 1984: 118). Herbst (1986: 105) also noted that 'the number of words used in LDOCE definitions could be estimated to lie between 5,000 and 10,000'. Hence, most learner's dictionaries are loath to measure the size of a CDV in terms of the number of senses of CDV items. That partly explains why the senses of CDV items are not differentiated and indicated in an English learner's dictionary.

Nevertheless, CIDE made some positive changes in the sense differentiation of polysemous CDV items. For instance, while all the CDV lists included the polysemous word *set*, only CIDE indicated that it was used in the meanings of [get ready] and [group], not in the senses of [position], [condition], [establish], [fix], etc. The sense coding system used in CIDE is called 'guide words'. Although OALD, LDOCE and MED have a similar sense coding system, respectively called 'short cuts', 'signposts' and 'menu', to guide the meanings of a polysemous headword, those dictionaries did not use the technique to differentiate the meanings of a CDV item.

5. POS indication of the LDV items

LDOCE introduced the policy that a POS label was given when there was a

restriction of some kind on a CDV item. For example, the POS label *n* was shown for the word *value*, indicating that in the definitions, the word was used only as a noun and not as a verb.

Yet, 'the parts of speech (POS) are not systematically mentioned' in the LDV list (Jansen, Mergeai and Vanadroye 1987: 83). Although the word *left* was included in the LDV list, some confusion over its POS would arise: it can be used in the definitions as an adjective, noun, adverb or past participle.

While the POS indication policy in the recent LDV lists did not show any change, it was noted that the Oxford 3000 labelled all its defining vocabulary items with the POS. Dictionary users will benefit from the clear indication of the POS of CDV items.

6. Affixes in the LDV lists

A derived word, such as *laziness*, though not on the LDV list, was still used in the definitions in LDOCE, for it can be formed by adding an affix to a CDV item. LDOCE included 54 affixes in the LDV1, 49 in the LDV2, and 30 respectively in the LDV3, LDV4 and LDV5.

The reason for the decrease of affixes in the LDV lists is that some affixes like *-al*, *-ly* and *-er* are too productive (Herbst 1986: 104). 'As a result of excessive reliance on affixation, some words like *free-dom* do not appear on the list. Others are: *busi-ness*, *for-th* and *un-less*.' (Jansen, Mergeai and Vanadroye 1987: 83) Although *independence* was not part of the LDV, it was still used in LDOCE3, for 'it can be constructed from its elements (**in-** + **depend** + **-ence**)' (Bogaards 1996: 289). That is why Herbst (1986: 114) argued that 'word formations whose meanings are not easily derivable ought to be listed as separate elements of the defining vocabulary'.

It was noted that the affix inclusion policy changed in the LDV3: 'the forms which are common, or which change their meaning when a prefix or suffix is added, (such as **acceptable** and **agreement**) are included in the full list.' The other English learner's dictionaries also limited their number of affixes in their CDVs. There are only 7 affixes in the Oxford 3000, namely 4 prefixes (*anti-*, *ex-*, *non-*, and *re-*), 1 suffix (*-ish*), and 2 combining forms (*mid-* and *self-*). Thus, some derivatives, such as *carefully*, *certainly/uncertain* and *encouragement*, were listed separately in the Oxford 3000. In the list of the Macmillan Defining Vocabulary (MDV), not a single affix was found. MED adhered to the principle that only inflected forms of the CDV items were used in the definitions. So some derivatives like *formally*, *improvement* and *leadership* became part of the MDV.

7. Multiword expression in the LDV lists

The role of multiword expressions in a CDV should not be underplayed, for a large proportion of language used in discourse is made up of formulaic se-

quences (Biber et al. 1999; Erman and Warren 2000; Wray 2002, 2008). Formulaic language such as phrasal verbs and collocations contributes to the naturalness of the defining language (Cowie 1999: 158). Phrasal verbs are, in particular, quite often used in a defining role (Cowie 1999: 111), but surprisingly there was only one phrasal verb in the LDV1 list, namely *wrap (up)* (Jansen, Mergeai and Vanadroye 1987: 83-84). LDOCE1 and LDOCE2 did not accord 'to each phrasal verb its full status as a separate lexical item in the defining vocabulary list' (Whitcut 1988: 52-53).

The recent editions of LDOCE paid more attention to the multiword expressions in the LDV. In terms of phrasal verbs, whereas there was only one in the LDV1, the number increased to 8 or 9 respectively in the LDV3, LDV4 and LDV5. Table 3 shows all the multiword expressions appearing in the LDV lists.

Table 3: Multiword expressions in the LDV lists

Types	LDV1	LDV2	LDV3	LDV4	LDV5
Phrasal verbs	<i>wrap (up)</i>		<i>deal with</i> <i>let go of</i> <i>lie down</i> <i>look after</i> <i>look for</i> <i>look sth up</i> <i>make into</i> <i>make up</i> <i>pick up</i>	<i>deal with</i> <i>let go of</i> <i>lie down</i> <i>look after</i> <i>look for</i> <i>look sth up</i> <i>make into</i> <i>make up</i> <i>pick up</i>	<i>deal with</i> <i>let go of</i> <i>lie down</i> <i>look after</i> <i>look for</i> <i>make into</i> <i>make up</i> <i>pick up</i>
Compounds	<i>postage stamp</i> <i>so-called</i> <i>upside-down</i>	<i>old-fashioned</i> <i>upside down</i>	<i>old-fashioned</i> <i>only just</i> <i>upside down</i>	<i>old-fashioned</i> <i>only just</i> <i>upside down</i>	<i>old-fashioned</i> <i>only just</i> <i>upside down</i>
Other multiword expressions	<i>all right</i> <i>according (to)</i> <i>no one</i> <i>owing to</i> <i>worthy (of)</i>	<i>according (to)</i> <i>no one</i> <i>owing to</i> <i>worthy (of)</i>	<i>according (to)</i> <i>as opposed to</i> <i>in spite of</i> <i>no one</i>	<i>according (to)</i> <i>as opposed to</i> <i>in spite of</i> <i>no one</i> <i>relating to</i>	<i>according (to)</i> <i>as opposed to</i> <i>in spite of</i> <i>no one</i> <i>relating to</i>

It is likely that the use of the multiword expressions in the definitions in LDOCE was not limited to those as shown in Table 3. To investigate the use of multiword expressions in the definitions, we compared those expressions in the LDV5 with those in the Oxford 3000. Among all the CDVs, the Oxford 3000 included the largest number of multiword expressions — altogether 126 ones (see Appendix 2). An overlap of 8 multiword expressions was found between the LDV5 and the Oxford 3000, namely *deal with*, *in spite of*, *look after*, *look for*, *make (sth) up*, *pick (sth) up*, *old-fashioned*, and *(as) opposed to*. Of the remaining 118 multiword expressions, it was found that 105 ones⁶ were actually used in LDOCE5. Only the following 13 were not used in the definitions in LDOCE5: *associated with*, *insist on*, *look forward to*, *rely on*, *go bad*, *on board*, *bound to*, *capable*

(of), a couple (of), for instance, on purpose, Yours sincerely, and Yours Truly. In other words, LDOCE5 used quite a large number of multiword expressions in the definitions, but did not acknowledge them in the LDV list.

It is not clear why LDOCE5 listed the multiword expressions as shown in Table 3, rather than the other ones that had been used. The LDV5 did include *consist, instead, rather, refer, relate, such*, etc., but when these items occurred in the definitions, they were, more often than not, used in the form of a set phrase, viz. *consist of, instead of, rather than, refer to, relate to, such as*, etc. It would be more helpful to learners if the fixed expressions were instead listed in the LDV.

In addition, the meanings of some multiword expressions are not transparent. Consider the meanings of *as well as* and *by accident* in the following definitions:

boarder *n* [C] 1 a student who stays at a school during the night, as well as during the day

bang² *v* 4 [T] to hit a part of your body, or something you are carrying, against something, by accident

The meanings of *as well as* and *by accident* can hardly be deduced from the meanings of their components *well* and *accident*. It is beyond the capabilities of some foreign learners to decode the meanings of such multiword expressions. Thus, it is a good idea to list them separately in the LDV and call learners' attention to them.

8. Conclusion

The recent LDV lists show some trends. Compared with the earlier versions of the LDV, the changes of CDV items in the recent lists are small. There is a relatively fixed number of LDV items. Over 85% of the LDV items fall into the ranges of the first three thousand most frequent words, and they should be familiar to ESL learners. For the defining purpose, some genus terms, grammatical terms as well as a few multiword expressions are also included in the LDV. The number of affixes in the LDV lists has been greatly reduced, and some common derivatives have been included in the full list.

There is still room for improvement in the LDV. It is insignificant to claim the use of only around 2,000 items in the definitions, for the actual number of the LDV items used in the definitions is much larger than was reported. The dictionary could indicate the senses, POS and multiword expressions of the LDV items more clearly. LDOCE can follow CIDE by using 'signposts' to show the senses of polysemous LDV items as used in definitions. The indication of POS of the LDV could be more consistent. Multiword expressions are indispensable to the definitions, and should be clearly indicated in the LDV list.

This article concerns CDVs in paper dictionaries. As we are entering a new era of electronic dictionaries, will CDVs still play a role? It is true that technical innovations offer the potential to extend a repertoire of modes of meaning representation from verbal explanation to multimodality, employing such devices as pictorial and animated illustrations, audio recordings and video clips (De Schryver 2003; Lew 2010). However, even in electronic dictionaries, definition still plays 'a prominent role' (Lew 2010: 292). 'Nor should the lexicographic definition contain words more difficult to understand than the explained word itself' (Zgusta 1971: 257). The basic principle of using simple words (i.e. CDV) in definitions is equally applicable to electronic dictionaries. Compared with printed dictionaries, one noticeable change in electronic dictionaries is that there might be more liberal use of words outside a CDV or a larger set of CDV items (cf. Lew 2010: 293). With mouse clicking or even hovering, hypertext links or pop-up windows in the electronic medium will provide instant access to the meaning of some words as used in definitions. Nevertheless, a caveat is that 'too many infrequent words in a definition may create comprehension problems' (Lew 2010: 293). Research issues, such as the pros and cons of a CDV in electronic dictionaries and the appropriate number of CDV items, deserve further investigation.

Acknowledgements

This study was supported by the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities in P.R. China (Project No. 12JJD740010), by the Programme for New Century Excellent Talents in University (NCET-11-0912), and by the Project of Humanities and Social Sciences at Universities in Guangdong Province (Project No. 11WYXM025). I am grateful to two anonymous reviewers for their constructive comments and to Ms Janice McAlpine, Prof. Chuming Wang, Prof. Jianhua Huang and Prof. Alison Wray for their encouragement.

Endnotes

1. The figures reported in this section are the count of word forms (including affixes) rather than lemmas. Therefore, *actor* and *actress* were treated as two word forms, though they were listed together in the LDV lists. Similarly, *according (to)*, *alcohol(ic)*, *arrangement(s)*, etc. were respectively regarded as two word forms. On the other hand, some phrases (e.g. *deal with*, *let go of*, *no one*), though being separated in space, were given full status as one word form. The homographs in the list, such as *bear (n.)* and *bear (v.)*, were respectively calculated.
2. The figures listed in Table 1 are the frequency of word types, not the frequency of tokens or word families (for the differences among the terms 'token', 'type' and 'word family', see Nation 2001: 7-8).

3. In Table 2, the senses of *note* were coded by the signposts as used in LDOCE5, and the sense ordering is identical with that of the dictionary.
4. In the interests of brevity and precision of a definition, almost all learner's dictionaries occasionally resort to some words outside a CDV list, and those words are printed in small capitals.
5. Another instance of usage of *command* in the definitions can be found, but it was not printed in small capitals:

processor *n* [C] 1 the central part of a computer that deals with the commands and information it is given.

As for non-LDV words in the definitions, there are some more typographical inconsistencies: *compound* is printed in small capitals in the definitions of **chloride**, **dioxide**, **halogen**, **halon**, **hydroxide**, **polymer**, **present participle**, **reagent**, **silica**, and **valence**, whereas in defining **hydrocarbon**, *compound* is in normal print; while *grave* in small capitals is used to define **archaeology**, **barrow**, **burial**, **bury**, **disinter**, **epitaph**, **ghoul**, **graveside**, **gravestone**, **headstone**, **tablet**, **tombstone**, **tumulus**, and **violate**, it is in normal print for the definition of **gravesite**. On the other hand, *physics*, part of the LDV5, has been carelessly set in small capitals in the following entries: **biophysics**, **Nobel Prize**, **physical**¹, **physical science**, **physicist**, **quantum**, **relativity**, **science**, and **solid-state**.

In addition, it is not clear why LDOCE4 and LDOCE5 removed from their CDV lists the items *compound* and *grave*, which respectively occurred 11 and 15 times in the definitions, while adding such an item as *physics*, which was used 13 times.
6. This includes some multiword expressions in small capitals (e.g. *credit card* and *make-up*).

References

A. Dictionaries

- CALD3 = Walter, E. (Ed.). 2008. *Cambridge Advanced Learner's Dictionary* (Third Edition). Cambridge: Cambridge University Press.
- CIDE = Procter, P. (Ed.). 1995. *Cambridge International Dictionary of English* (First Edition). Cambridge: Cambridge University Press.
- COBUILD1 = Sinclair, J. (Ed.). 1987. *Collins COBUILD English Language Dictionary* (First Edition). London/Glasgow: Collins.
- COBUILD2 = Sinclair, J. (Ed.). 1995. *Collins COBUILD English Dictionary* (Second Edition). London/Glasgow: HarperCollins.
- COBUILD6 = Sinclair, J. (Ed.). 2008. *Collins COBUILD Advanced Learner's English Dictionary* (Sixth Edition). Glasgow: HarperCollins.
- CULD = Kirkpatrick, E.M. (Ed.). 1980. *Chambers Universal Learners' Dictionary*. Edinburgh: Chambers.
- LDOCE1 = Procter, P. (Ed.). 1978. *Longman Dictionary of Contemporary English* (First Edition). Harlow: Longman.
- LDOCE2 = Summers, D. (Ed.). 1987. *Longman Dictionary of Contemporary English* (Second Edition). Harlow: Longman.

- LDOCE3 = Summers, D. (Ed.). 1995. *Longman Dictionary of Contemporary English* (Third Edition). Harlow: Longman.
- LDOCE4 = Summers, D. (Ed.). 2003. *Longman Dictionary of Contemporary English* (Fourth Edition). Harlow: Pearson Education.
- LDOCE5 = Mayor, M. (Ed.). 2009. *Longman Dictionary of Contemporary English* (Fifth Edition). Harlow: Pearson Education.
- MED1 = Rundell, M. (Ed.). 2002. *Macmillan English Dictionary for Advanced Learners* (First edition). Oxford: Macmillan Education.
- MED2 = Rundell, M. (Ed.). 2007. *Macmillan English Dictionary for Advanced Learners* (Second Edition). Oxford: Macmillan Education.
- MW11 = Mish, F. (Ed.). 2003. *Merriam-Webster's Collegiate Dictionary* (Eleventh Edition). Springfield, Massachusetts: Merriam-Webster.
- NMED = West, M.P. and J.G. Endicott. (Eds.). 1935. *The New Method English Dictionary*. London: Longmans, Green.
- OALD3 = Hornby, A.S., A.P. Cowie and L.J. Windsor. (Eds.). 1974. *Oxford Advanced Learner's Dictionary of Current English* (Third Edition). Oxford: Oxford University Press.
- OALD4 = Cowie, A.P. (Ed.) 1989. *Oxford Advanced Learner's Dictionary of Current English* (Fourth Edition). Oxford: Oxford University Press.
- OALD5 = Crowther, J. (Ed.). 1995. *Oxford Advanced Learner's Dictionary of Current English* (Fifth Edition). Oxford: Oxford University Press.
- OALD6 = Wehmeier, S. (Ed.). 2000. *Oxford Advanced Learner's Dictionary of Current English* (Sixth Edition). Oxford: Oxford University Press.
- OALD7 = Wehmeier, S. (Ed.). 2005. *Oxford Advanced Learner's Dictionary of Current English* (Seventh Edition). Oxford: Oxford University Press.
- OALD8 = Turnbull, J. (Ed.). 2010. *Oxford Advanced Learner's Dictionary of Current English* (Eighth Edition). Oxford: Oxford University Press.
- WNWD2 = Guralnik, D.B. (Ed.). 1980. *Webster's New World Dictionary of the American Language* (Second College Edition). New York: Simon and Schuster.

B. Other literature

- Béjoint, H.** 2000. *Modern Lexicography: An Introduction*. Oxford/New York: Oxford University Press.
- Biber, D. et al.** 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Bogaards, P.** 1996. Dictionaries for Learners of English. *International Journal of Lexicography* 9(4): 277-320.
- Bogaards, P.** 2003. MEDAL: A Fifth Dictionary for Learners of English. *International Journal of Lexicography* 16(1): 43-55.
- Bogaards, P.** 2008. Frequency in Learners' Dictionaries. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15-19 July 2008*: 1231-1236. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Bullon, S. and G. Leech.** 2007. Longman Communication 3000 and the Longman Defining Vocabulary. Bullon, S. and G. Leech (Eds.). 2007. *Longman Communication 3000*: 1-7. Harlow: Pearson Education.

- Clark, G. 2003. Analyzing Word Meaning and Representation by Navigating Dictionary Definitions. <http://www.crsc.uqam.ca/en/SiteGraham/index.html>.
- Cowie, A.P. 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Oxford University Press/Clarendon Press.
- De Schryver, G-M. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Erman, B. and B. Warren. 2000. The Idiom Principle and the Open-choice Principle. *Text: An Interdisciplinary Journal for the Study of Discourse* 20(1): 29-62.
- Fontenelle, T. 2009. Linguistic Research and Learner's Dictionaries: the *Longman Dictionary of Contemporary English*. Cowie, A.P. (Ed.). 2009. *The Oxford History of English Lexicography. Volume II: Specialized Dictionaries*: 412-435. Oxford: Oxford University Press.
- Fox, G. 1989. A Vocabulary for Writing Dictionaries. Tickoo, M.L. (Ed.). 1989. *Learners' Dictionaries: State of the Art*: 153-171. Singapore: SEAMEO Regional Language Centre.
- Heatley, A., I.S.P. Nation and A. Coxhead. 2002. RANGE and FREQUENCY Programs. <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>.
- Herbst, T. 1986. Defining with a Controlled Defining Vocabulary in Foreign Learners' Dictionaries. *Lexicographica: International Annual for Lexicography* 2: 101-119.
- Herbst, T. 1996. On the Way to the Perfect Learners' Dictionary: A First Comparison of OALD5, LDOCE3, COBUILD2 and CIDE. *International Journal of Lexicography* 9(4): 321-357.
- Jansen, J., J.P. Mergeai and J. Vanadroye. 1987. Controlling LDOCE's Controlled Vocabulary. Cowie, A.P. (Ed.). 1987. *The Dictionary and the Language Learner. Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985*. *Lexicographica. Series Maior* 17: 78-94. Tübingen: Max Niemeyer.
- Lew, R. 2010. Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries. *Lexikos* 20: 290-306.
- MacFarquhar, P.D. and J.C. Richards. 1983. On Dictionaries and Definitions. *RELC Journal* 14(1): 111-124.
- McCreary, D.R. and E. Amacker. 2006. Experimental Research on College Students' Usage of Two Dictionaries: A Comparison of the *Merriam-Webster Collegiate Dictionary* and the *Macmillan English Dictionary for Advanced Learners*. Corino, E., C. Marello and C. Onesti (Eds.). 2006. *Proceedings XII EURALEX International Congress, Turin, Italy, September 6th-9th, 2006*: 871-885. Alessandria: Edizioni dell'Orso.
- Michels, A. and J. Noël. 1984. The Pro's and Con's of a Controlled Defining Vocabulary in a Learner's Dictionary. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*. *Lexicographica. Series Maior* 1: 385-394. Tübingen: Max Niemeyer.
- Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Neubauer, F. 1984. The Language of Explanation in Monolingual Dictionaries. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83. Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*. *Lexicographica. Series Maior* 1: 117-123. Tübingen: Max Niemeyer.
- Stein, G. 1979. The Best of British and American Lexicography. *Dictionaries: Journal of the Dictionary Society of North America* 1: 1-23.

- Whitcut, J.** 1988. Lexicography in Simple Language. *International Journal of Lexicography* 1(1): 49-55.
- Wray, A.** 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wray, A.** 2008. *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton.

Appendix 1: The LDV items outside the 3,000 most frequent words

abbreviation, absent, adjective, adverb, -al, alphabet, -ance, anxiety, anxious, -ation, atom, beak, beneath, biology, broadcast, bubble, bullet, carriage, cattle, ceremony, chemistry, chin, clay, comb, consist, corn, creature, criticize, curve, decay, deceive, decrease, defeat, dis-, dismiss, eager, email, -ence, enclose, -er, explosion, explosive, fasten, fever, flesh, flour, former, -ful, girlfriend, goat, herb, hollow, horizontal, horn, -ic, -ical, im-, importance, in-, -ing, ink, insect, inwards, -ion, ir-, -ish, -ity, -ive, -ize, kilogram, kilometre, kiss, kneel, laughter, leather, lion, loyal, loyalty, -ly, -ment, military, mineral, mist, monkey, navy, -ness, network, non-, noun, obey, ocean, old-fashioned, onion, opponent, oxygen, parallel, participle, pepper, phrase, physics, plural, poet, poetry, preserve, president, priest, profession, pronunciation, protest, quarrel, re-, rice, romantic, sauce, scatter, scissors, self-, sideways, singular, skirt, slippery, slope, snake, soap, software, sour, spacecraft, spice, spicy, stem, sting, swell, sword, thirsty, thread, tobacco, tomato, transparent, tribe, tropical, ugly, un-, underwear, universe, upright, valley, verb, vertical, vowel, waist, wealth, weave, website, wheat (154)

Appendix 2: Multiword expressions in the Oxford 3000

Phrasal verbs: *approve (of), associated with, based on, give birth (to), take care (of), care for, deal with, depend (on), disapprove (of), fall over, find out (sth), make fun of, get on, get off, give sth away, give sth out, give sth up, go down, go up, involved in, insist (on), leave out, look after, look at, look for, look forward to, make sth up, take notice of, pick sth up, put sth on, put sth out, refer to, relate (to), rely on, get rid of, set fire to sth, sit down, stand up, stick out, switch off, switch on, take sth off, take (sth) over, take part (in), throw sth away, tie sth up, wake (up), wind sth up (48)*

Compounds: *credit card, ice cream, well known, make-up, old-fashioned, post office, prime minister, swimming pool (8)*

Other multiword expressions: *one another, by accident, take action, in addition (to), in advance, all right, apart from, fall asleep, pay attention, go bad, on behalf of /on sb's behalf, a bit, on board, be born, bound to, be called, capable (of), in case (of), in charge of, in common, consist of, in control of, under control, a couple, of course, in detail, in the end, in exchange (for), fall asleep, in favour/favor (of), a few, at first, in front of, in general, be going to, good at, good for, have to, in honour/honor of, in a hurry, for instance, instead of, at least, a little, a lot (of)/lots (of), by means of, in memory of, next to, opposed to, in order to, out (of), in public, on purpose, rather than, the rest, be sick, feel sick, Yours sincerely, so that, as soon as, in spite of, such as, make sure, take place, thank you, Yours Truly, used to sth/doing sth, used to, as well (as), go wrong (70)*