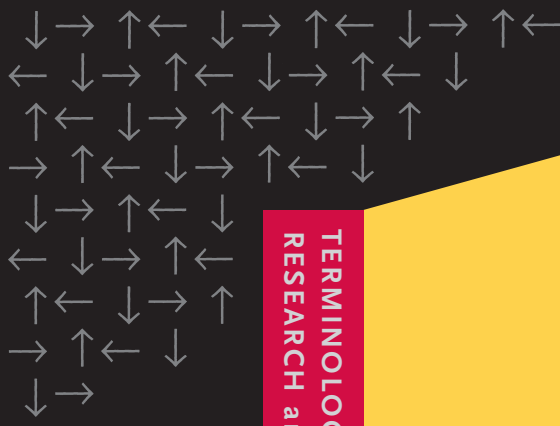


Copyright 2021. John Benjamins Publishing Company. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair use permitted under U.S. or applicable copyright law.



TERMINOLOGY and LEXICOGRAPHY
RESEARCH and PRACTICE 22

Defining with Simple Vocabulary in English Dictionaries

Mariusz Piotr Kamiński

John Benjamins Publishing Company

Defining with Simple Vocabulary in English Dictionaries

Terminology and Lexicography Research and Practice (TLRP)

ISSN 1388-8455

Terminology and Lexicography Research and Practice aims to provide in-depth studies and background information pertaining to Lexicography and Terminology. General works include philosophical, historical, theoretical, computational and cognitive approaches. Other works focus on structures for purpose- and domain-specific compilation (LSP), dictionary design, and training. The series includes monographs, state-of-the-art volumes and course books in the English language.

For an overview of all books published in this series, please see benjamins.com/catalog/tlrp

Editors

Marie-Claude L'Homme
University of Montreal

Kyo Kageura
University of Tokyo

Volume 22

Defining with Simple Vocabulary in English Dictionaries
by Mariusz Piotr Kamiński

Defining with Simple Vocabulary in English Dictionaries

Mariusz Piotr Kamiński

University of Applied Sciences in Nysa

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/tlrp.22

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2021003374 (PRINT) / 2021003375 (E-BOOK)

ISBN 978 90 272 0859 0 (HB)

ISBN 978 90 272 6000 0 (E-BOOK)

© 2021 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Company · <https://benjamins.com>

Table of Contents

List of Tables	IX
List of Figures	XI
Abbreviations used in the text	XIII
Acknowledgements	XV
Introduction	1
1. Historical background and issues	1
2. Research questions	5
3. Studies presented in the book	7

CHAPTER 1

Pioneering research on vocabulary control	11
1.1 Early interests in vocabulary control	11
1.2 The objective approach to vocabulary selection	13
1.2.1 Edward L. Thorndike	14
1.2.2 Ernest Horn	15
1.2.3 Lawrence Faucett and Itsu Maki	15
1.2.4 The strengths and limitations of statistical word-counts	16
1.3 The subjective approach	18
1.3.1 Harold Palmer	19
1.3.2 Michael West	23
1.4 The logical approach	28
1.5 The General Service List: Confrontation of the competing approaches	33
1.5.1 The Carnegie Conferences	33
1.5.2 The General Service List	34
1.5.3 Criticism of the GSL	42
1.6 Summary	45

CHAPTER 2

Other major projects on vocabulary control	47
2.1 Word-lists for pedagogical purposes	47
2.1.1 The American Heritage Word Frequency Book (1971)	48
2.1.2 The Threshold Level English (1975)	48
2.1.3 The Cambridge English Lexicon (1980)	49
2.1.4 Word-lists from popular corpora: the Brown (1967), the LOB (1978), the Bank of English (1991), and the BNC (1990s)	50
2.1.5 The University Word List (1984)	53
2.1.6 The Academic Word List (2000)	55
2.1.7 The Common Core Vocabulary (2002)	56
2.1.8 A New Academic Vocabulary List (2013)	57
2.1.9 A New General Service List (2013)	58
2.2 The GSL against other word-lists and corpora	59
2.3 Controlled vocabularies for professional and human-computer communications	62
2.3.1 Controlled English for writing service manuals	62
2.3.2 Text production for machine translation	66
2.4 Identifying a core vocabulary	69
2.5 Wierzbicka's Natural Semantic Metalanguage and relevant projects	78
2.5.1 Natural Semantic Metalanguage	78
2.5.2 A primitive-based dictionary	82
2.5.3 Minimal English	85
2.5.4 Defining with primitives for cross-linguistic comparison	85
2.6 Summary	87

CHAPTER 3

Defining with a controlled vocabulary: Preliminary insights	89
3.1 What is the dictionary definition?	90
3.2 Defining in plain English: A historical overview of native-speaker dictionary tradition	93
3.3 Defining in learners' dictionaries between 1935 and 1986	104
3.3.1 Early learners' dictionaries: NMED and ISED	105
3.3.2 Revisions of ALD and emergence of new competitors: 1963–1986	109
3.3.3 Conclusions	117
3.4 Review of previous studies of defining vocabularies	117
3.4.1 Summary of previous studies and implications for the current research	130
3.4.2 Research hypotheses	134

CHAPTER 4

Quantitative analyses	137
4.1 Research design	137
4.1.1 Materials	137
4.1.2 Data selection	138
4.1.3 Research tools and reference lists	142
4.1.4 Data analysis	145
4.2 Analysis of receptive vocabulary load and other quantitative comparisons	148
4.2.1 Receptive vocabulary load of definitions	148
4.2.2 Receptive vocabulary load of DV lists	152
4.3 Similarity of content	157
4.3.1 Size of the lists	157
4.3.2 HCA: Similarities between the lists	159
4.3.3 Overlapping and unique items	160
4.3.4 The Core Defining Vocabulary	163
4.3.5 Multiword items	166
4.3.6 Word-building elements	167
4.4 Conclusions	168

CHAPTER 5

Qualitative analysis	171
5.1 Research materials and methodology	171
5.1.1 Materials	171
5.1.2 Data selection	172
5.2 Learners' dictionaries (1987–2015)	175
5.2.1 LDOCE2 (1987)–LDOCE6 (2014)	175
5.2.2 OALD4 (1989) – OALD9 (2015)	185
5.2.2.1 The DV of OALD5 (1995) and subsequent editions	187
5.2.3 COBUILD1 (1987) – COBUILD8 (2014)	196
5.2.4 CALD1 – CALD4 (1995–2013)	202
5.2.5 MEDAL (2002–2007)	207
5.2.6 MWALED (2008)	211
5.2.7 DV in digital dictionaries	213
5.2.8 Summary and conclusion	215
5.3 Native-speaker dictionaries (1984–2011)	217
5.3.1 Analysis	218
5.3.2 Conclusions	225

CHAPTER 6

The defining vocabulary and the user	227
6.1 Design of the experiment	227
6.1.1 Subjects	228
6.1.2 Materials	228
6.1.3 Procedure	231
6.2 Results	232
6.2.1 Comprehension of dictionary definitions	232
6.2.2 Comprehension of invented definitions	235
6.2.3 Other factors affecting comprehension score	237
6.2.4 Perception of definitions	240
6.2.5 Summary of the results	241
6.3 Discussion	241
6.4 Conclusions	244
General conclusions	247
1. Main findings	247
2. Implications for definition writing and DV design	249
3. Further research	252
APPENDIX 1	
The Core Defining Vocabulary (CDV)	253
APPENDIX 2	
Word Familiarity Page	263
Task 1, version 1 (with <i>LDOCE6</i> definitions)	264
Task 1, version 2 (<i>COBUILD-8</i>)	268
Task 1, version 3 (<i>ODE-2</i>)	272
Task 1, version 4 (<i>Chambers-12</i>)	277
Task 2, version 1 (all the dictionaries)	281
Task 2, version 2 (all the dictionaries)	288
References	297
Dictionaries	297
Printed dictionaries	297
Dictionaries on CD-ROM/DVDs	300
Online dictionaries	301
Other literature	301
Index	325

List of Tables

Table 1. Input for hierarchical cluster analysis (part of the data).	147
Table 2. Cumulative percentage coverage of definitions in learners' dictionaries by Base Lists	150
Table 3. Cumulative percentage coverage of definitions in native-speaker dictionaries by Base Lists	154
Table 4. Cumulative percentage coverage of DV lists by Base Lists (affixes excluded).	156
Table 5. Overlapping and unique vocabulary items in the early DVs.	160
Table 6. Overlapping and unique vocabulary items in the recent DVs.	161
Table 7. Overlapping and unique vocabulary items in the <i>OALD</i> DVs.	161
Table 8. Overlapping and unique vocabulary items in the <i>LDOCE</i> DVs.	161
Table 9. Overlapping and unique vocabulary items in the <i>CALD</i> DVs	161
Table 10. Overlapping and unique vocabulary items in the <i>MEDAL</i> DVs	161
Table 11. Examples of vocabulary items unique to a particular DV list.	162
Table 12. Multiword items in the DVs of the early and recent editions of the dictionaries.	167
Table 13. Word-building elements in the DVs.	168
Table 14. Headwords selected for the qualitative research.	174
Table 15. Students' familiarity with headwords.	233
Table 16. Comprehension scores relative to the approach to defining vocabulary and RVL (total N = 215).	234
Table 17. Percentages of correct responses.	235
Table 18. Students' familiarity with the words.	236
Table 19. The data in comprehension task with invented definitions (total N = 215).	236

Table 20. Percentages of correct responses.	237
Table 21. Percentage of high-frequency and wide-range words, and comprehension of the definitions for headwords 1–10.	238
Table 22. Percentage of synonym definitions, definition length, and comprehension of definitions for headwords 1–10.	239
Table 23. The data in perception task (N = 250).	240

List of Figures

Figure 1. Murray's vision of the English vocabulary (from Murray 1888, xvii).	70
Figure 2. The size of the DV lists in the earliest and the latest editions published in 1987–2015.	158
Figure 3. Hierarchical cluster analysis: Visualization of similarity of DV lists.	159
Figure 4. Comprehension scores by dictionary.	234
Figure 5. Correlation between comprehension score and receptive vocabulary load (RVL).	234
Figure 6. Comprehension of invented definitions.	236
Figure 7. Correlation between comprehension score and the percentage of high-frequency and wide-range words in definitions.	238
Figure 8. Correlation between comprehension score and the percentage of synonym definitions.	239
Figure 9. Correlation between comprehension score and average definition length.	239
Figure 10. Students' perception of definition helpfulness.	240

Abbreviations used in the text¹

AVL	New Academic Vocabulary List
AWL	Academic Word List
BNC	British National Corpus
CCV	Common Core Vocabulary
CEFR	Common European Framework of Reference for Languages
COCA	Corpus of Contemporary American English
DV	defining vocabulary
EFL	English as a Foreign Language
ESL	English as a Second Language
FL	foreign language
GSL	General Service List
H	research hypothesis
IPA	International Phonetic Alphabet
IRET	Institute for Research in English Teaching
L1	learner's mother tongue
L2	learner's foreign language
LOB	Lancaster-Oslo-Bergen corpus
MT	machine translation
NGSL	New General Service List
NLP	natural language processing
NSM	Natural Semantic Metalanguage
PACE	Perkins Approved Clear English
POS	part of speech
RQ	research question
RVL	receptive vocabulary load
STE	Simplified Technical English
ukWaC	the corpus of British English compiled from the .uk domain
UWL	University Word List
WSC	the Wellington Corpus of Spoken English (New Zealand)
WSD	word sense disambiguation

1. Abbreviations for dictionary titles are listed under References

Acknowledgements

I am pleased to acknowledge the helpful feedback from several researchers, lexicographers, publishers, and librarians. I owe a special debt of gratitude to Professor Tadeusz Piotrowski for inspiration, encouragement, and suggestions. I am grateful to Michael Rundell, the Editor-in-Chief of Macmillan Dictionaries, for sharing with me valuable insights into the defining policy of *MEDAL* and for enlightening my understanding of lexicographic practice in other dictionaries. Thanks are also due to publishers who provided me with the defining vocabulary lists, in particular Wendalyn Nichols, publishing manager from Cambridge University Press, and Kateryna Prokhorova from Pearson Technical Product Support. I also thank Professor Hai Xu at the Guangdong University of Foreign Studies, China, for sharing with me a few electronic versions of vocabulary lists. I would also like to express my thanks to the team of Collins Dictionaries at HarperCollins for supplying me with clues about the *COBUILD* defining policy.

A large part of the quantitative research reported here has been conducted with the AntWordProfiler program developed by Professor Laurence Anthony of Waseda University, and the corpus database compiled by Professor Paul Nation of Victoria University of Wellington, to whom I acknowledge my gratitude.

In the preparation of this work I made extensive use of several libraries in England and Poland. The research trip to the Bodleian Library, Oxford, was funded by the European Society for the Study of English (ESSE) post-doc bursary (type B) awarded to me in 2019. I am much indebted to the members of the library staff at the University of Applied Sciences, Nysa, in particular Bogumiła Wojciechowska-Marek, director of the library, and Krzysztof Balcerek, an inter-library loan officer, for their time and assistance in making research materials available to me.

Special thanks are due to Dr Elżbieta Szymańska-Czaplak, PhD, Dean of the Faculty of English Philology at University of Opole, for granting me the permission to conduct the investigation in Opole. The students of the University of Opole and of the University of Applied Sciences, Nysa, were absolutely invaluable for this study. The completion of this book would not have been possible without their participation in the experiment.

Finally, I would like to thank two anonymous reviewers and the series editors for their detailed comments on an earlier version of this book. Their critical and insightful remarks greatly contributed to the final shape of the book. Needless to say, any mistakes and inaccuracies are my own.

.

Introduction

1. Historical background and issues

In the 1899 account of principles for constructing dictionaries, the eminent philologist Henry Sweet remarked that “[t]he first business of a dictionary is to give the meanings of the words in plain, simple, unambiguous language” (1899, 264). Sweet made this suggestion in the context of general dictionaries for native speakers, but it also applies to dictionaries for foreign learners.

Dictionaries are invaluable tools for their users, providing them with perhaps the most explicit account of word meanings and uses (Stein 2002a). Advantages of using dictionaries are obvious to those readers who encounter an unfamiliar word in a text being read (Hulstijn 1993; Bogaards 1998, 2003; Atkins and Rundell 2008). In order to overcome the problem, the reader may attempt to guess the meaning from context and continue reading, but if the quandary persists, they often decide to turn to a dictionary for help. It is a typical function of the dictionary to aid the reader, whether a native speaker or a foreign language (FL) learner, in understanding texts (Barnhart 1962; Quirk 1973; Tomaszczyk 1979; Greenbaum et al. 1984; Summers 1988; Piotrowski 2001). Strictly speaking, the dictionary is also a text, albeit an untypical one. It is an interrupted text intended for quick reference rather than extended reading from cover to cover (Miodunka 1989; Chlebda 2010). Whether the reader succeeds in solving the comprehension problem depends on a number of factors. If the dictionary at hand is monolingual, success depends, *inter alia*, on whether the definition of the word being looked up is intelligible to the user (Atkins and Rundell 2008, 412).

It can be argued that definitions that make use of rare and specialized words pose a greater challenge for readers than those which dispense with such vocabulary items. Such definitions are likely to become circular for learners. Michael West, a founding father of English as a Foreign Language (EFL) lexicography and a key figure of the vocabulary control movement (see 1.3.2), considered the selection of words for a definition vocabulary as “the first and most fundamental problem” for compilers of learners’ dictionaries (West 1935, 7). As he remarked, “If you are writing the explanations of English words in English, how do you know that the foreigner will know the explaining words which you use?” (West, 7).

West pioneered the use of a restricted defining vocabulary (DV) in the *New Method English Dictionary* (NMED) in 1935. In this dictionary, he wrote definitions using a limited vocabulary of 1,490 items. A large number of the items had been selected prior to defining, though the content of the vocabulary was revised as the dictionary progressed. Restricting a definition vocabulary to a specified subset of the lexicon was a radical departure from the established tradition of native-speaker dictionaries, which used no such constraints (Rundell 2008a). West's approach to writing definitions was intended to make them accessible to learners with relatively low language proficiency.

A. S. Hornby adopted a different approach to definitions in the *Idiomatic and Syntactic English Dictionary* (ISED), published in 1942. He used no special DV of a kind pioneered by West, but whenever possible he strictly adhered to the fundamental lexicographic principle that a word should be defined by simpler words (see 3.2.1.2). Following Whitcut (1986), one may say that Hornby's definition vocabulary was merely controlled to some extent, rather than restricted to a particular set of words. Although Hornby's approach presupposes a certain degree of control of definition vocabulary, I will follow the met-alexicographic convention and reserve the qualifiers *controlled*, *restricted*, and *limited* for definitions written with a restricted set of vocabulary items (West's approach), and use the words *unrestricted* or *uncontrolled* to refer to definitions following Hornby's tradition.

Both NMED and ISED served as models for subsequent dictionaries for learners. For example, all the editions of *Longman Dictionary of Contemporary English* (LDOCE) follow West's approach, specifying which words are allowable before the lexicographers start writing definitions. In turn, all the editions of *Collins COBUILD Advanced Learner's Dictionary* (COBUILD) conform to Hornby's tradition. Although this dictionary dispenses with an explicit list, the lexicographers are responsible for monitoring what goes into definitions and paying attention to avoid defining *obscurum per obscurius* (Fox 1989). As we shall see (3.4), the two approaches are not accepted indiscriminately by met-alexicographers, and both have supporters and critics. For example, Whitcut (1986, 118) believes that intelligibility is best achieved by using a restricted defining vocabulary, whereas Fox (1989: 156) argues against this approach because it is hard to predict in advance which vocabulary items are necessary for definitions.

Restricted defining vocabularies have their strengths and weaknesses. For example, researchers argue that they lead to definitions being oversimplified, imprecise and vague (e.g., Stein 1979; Lew 2010). Fox (1989) claims that such definitions fail to represent natural language. On the other hand, in several studies of dictionaries published up to 1995, researchers report that learners judge definitions couched in a restricted defining vocabulary (as in LDOCE) as being easier to

understand than definitions written with no such restrictions (MacFarquhar and Richards 1983; Herbst 1986; Nesi 1998). With the emergence of new dictionaries and editions in the second half of the 20th century, DVs were subjected to revision. Restricted defining vocabularies are now used by several major dictionaries, but there is a considerable disagreement as to which vocabulary items they should consist of (4.3.3).

As mentioned earlier, West (1935, 7) raised the problem of how lexicographers know that the words used in their definitions are known to the learner. He added that the same problem arises for compilers of dictionaries for native speakers. Whitcut (1986) argues that some degree of vocabulary control is necessary for the latter dictionaries. Yet even a cursory look at native-speaker dictionaries published up to 1980s reveals that the basic lexicographic principle of defining words in intelligible terms was often ignored (see 3.2). It seems, however, that the more recent dictionaries for native speakers are more careful about selecting words for their definitions.

Some researchers observe in recent decades a trend in this lexicographic genre to make information easily accessible to the user, a development in which EFL dictionaries have blazed a trail. Atkins and Rundell (2008, 432) report a “gradual move away from conventional models towards styles that are both more explicit and more user-friendly”. To illustrate this finding, they quote a definition of **decadent** in *Collins English Dictionary (CoED)* and contrast it with the corresponding definition from the *Collegiate*⁸:

decadent marked by decay or decline (Collegiate⁸ 1973)

decadent characterized by decay or decline, as in being self-indulgent or morally corrupt (CoED 1979)

The latter definition improves on the former by dispensing with a lexicographic formula “marked by”, and adding extra clues as to the typical context of use of the word being defined (Atkins and Rundell 2008). As can be seen, the changes are not necessarily tantamount to greater restrictions on the choice of the defining vocabulary (note the addition of rare *self-indulgent*). However, the replacement of *marked (by)*, which is used in a non-central meaning (e.g. *LDOCE4* records it as the 7th sense in the entry), with a more transparent *characterized (by)* leaves little doubt as to the meaning of the substituting verb even for less proficient users. Another telling example is reported by Kamiński (2013), who finds that *Chambers*⁸ (1993) makes it easier for (less proficient) users to understand the definition of **madcap** by replacing the rare word *frolicsome* with the more common one *playful*, while retaining the same syntax of the definition:

madcap an exuberantly frolicsome person (Chambers⁷)

madcap an exuberantly playful person (Chambers⁸)

It would be interesting to explore whether such changes are implemented in a regular or merely haphazard manner. In changes towards greater accessibility of definitions, it is only natural to expect a certain degree of control of the vocabulary of definitions. This may be manifested, *inter alia*, by preference for more common vocabulary items over those which are rare, technical, or specialized in some way. If it is really the case, then one can argue that native-speaker dictionaries follow learners' dictionaries at least to some extent. After all, it is not unusual for two distinct lexicographic traditions to draw innovations from each other.

It was Randolph Quirk who in 1986 envisaged that some of the innovations which were "first implemented to the advantage of EFL dictionaries" would "subsequently be incorporated into the mainstream tradition" of native-speaker dictionaries (Quirk 1986, 4). One of the reasons is commercial and profit-driven: "any innovation which looks likely to improve the product is sooner or later adopted by" other dictionaries (Rundell 2006, 739; Rundell 2008b, 198) as long as they meet the prospective users' needs (see also Kernerman 1996, 408; Dziemianko and Lew 2006; Piotrowski 2009). Some novelties have already leaked into native-speaker dictionaries. One of the most noticeable signs of this impact is the indication of pronunciation in the International Phonetic Alphabet (IPA), which has dominated almost entirely British monolingual dictionaries (Landau 2001). For example, *COD* introduced IPA transcription in 1990, breaking with the long-established tradition of respelling in this dictionary (Kamińska 2014). Another area of influence is usage labels, where *Collins English Dictionary* (1979), *Reader's Digest Great Illustrated Dictionary* (1984) and *Longman Dictionary of the English Language* (1984) follow earlier EFL dictionaries in marking formality of some words and expressions (Ilson 1990). We will see in Section 5.3 how recent native-speaker dictionaries benefit from learners' dictionaries to make their definitions more accessible to users.

The relationship between the two lexicographic genres is by no means unidirectional; suffice it to say that they mutually influence each other and that early learners' dictionaries are lineal descendants of native-speaker dictionaries from the early decades of the 20th century (see 3.2). Although the former inherited many qualities from the latter, learners' dictionaries gradually developed their own distinctive characteristics, which were inspired by concern for the different needs of the target readership (McArthur 1986). Thus, along with a defining vocabulary, they developed strategies which broke with the dense and formulaic style of classical definitions; among these strategies were full-sentence definitions and folk-defining techniques which are believed to make definitions user-friendly (Adamska-Sałaciak 2012). These features were driven by the necessity to make information more accessible to the learner, and by the desire to place the

meaning of a word being explained in a typical context (Hanks 1987, 117; Hanks 1990; Atkins and Rundell 2008, 433). Other distinctive characteristics of learners' dictionaries include complementation patterns, emphasis on collocations, pronunciation notation in IPA, and a more developed system of labels informing the learner on the nuances of meaning and usage.

This book highlights the fact that the dictionary is an important tool in language pedagogy, but its full potential can be exploited if definitions are intelligible to learners. The dictionary not only aids the learner in overcoming comprehension problems as they arise while reading a text, but also contributes to them learning the words they have looked up (Laufer 1993; Nesi 2000; Ronald 2002; Abecassis 2007). Just as low-frequency words that keep recurring in texts are learnt through context (Jenkins et al. 1984; Horst et al. 1998; Waring and Takaki 2003; Webb 2007), so repeated consultation of their definitions helps learning the words. As Béjoint (1987, 104) remarks, "frequent and careful consultation of the dictionary can lead to a better command of the language". Moreover, language learners who have access to definitions of unfamiliar words not only perform better on comprehension tests than those who do not have definitions at their disposal (Black 1986; Luppescu and Day 1993), but they also remember the words they have looked up for a relatively long time (Bogaards 1991).

Although dictionaries are primarily designed to solve immediate problems encountered by language learners rather than to help them expand their vocabulary in the long term (Lew 2012), the latter still appears to be a promising potential of dictionary consultation. This is an argument in favor of using dictionaries, which becomes even more powerful if one realizes the fact that "vocabulary acquisition has been identified as the most serious learning problem" for foreign language learners (Stein 2002b, 49). This fact is noteworthy, as anybody trying to learn a foreign language knows that vocabulary plays a crucial role in communication. However, the success of vocabulary development through repeated definition look-up can only be achieved when definitions are comprehensible to learners.

2. Research questions

The present book has three major goals. The first is to emphasize the role played by vocabulary control in definitions in learners' dictionaries. Second, the book aims to examine lexicographic strategies for controlling the vocabulary of definitions, and third, to explore their implications on definition writing in learners' dictionaries. The studies presented in the book address several research questions, which are listed below in the order they are dealt with in the text. I will elaborate on them in Section 3.4.1 following a review of previous studies.

RQ1. Do definitions written with a different approach to defining vocabulary (with and without an explicit set of DV items) present different levels of vocabulary burden for their users? In other words, do the vocabularies used in these definitions impose the same levels of difficulty on learners? A similar question arises with respect to native-speaker dictionaries: to what extent their definitions differ in this regard? Native-speaker dictionaries will be explored in an attempt to evaluate them as potential reference sources for learners. The idea that learners may benefit from such dictionaries was inspired by the aforementioned reports on definitions employing more user-friendly styles than conventional models (Atkins and Rundell 2008; Kamiński 2013).

RQ2. If learner's dictionaries disagree on the selection of defining vocabulary items, then to what extent their DVs differ and to what extent they overlap? Is there a core defining vocabulary that is most useful for learners in understanding dictionary definitions?

RQ3. Considering criticism of restricted DVs, one may expect that contemporary dictionaries have made improvements to their defining policies. Are recent dictionaries more consistent in controlling their definition vocabularies than earlier works? This question applies to both learner's and native-speaker dictionaries.

RQ4. A major concern of this book is the study of the contribution of a controlled vocabulary to the learner's understanding of definitions. Cowie (1989, 47) notes that a defining vocabulary is believed to "facilitate understanding of unfamiliar words by defining them in terms of words which the dictionary user already knows", but he hastens to add that it is not tantamount to saying that the vocabulary "provides a fuller or more precise understanding of those words". Thus, Cowie asks, "[h]ow far then is the principal aim of the CV [i.e. controlled vocabulary] realized?" (1989, 47). The current research attempts to address this question by examining learners' performance in comprehension tasks with a dictionary definition as an aid. The question of our interest is to what extent definitions written with a controlled vocabulary are successful in explicating word meanings. In order to eliminate the effect of external factors, one part of this study will focus on invented rather than authentic dictionary definitions.

RQ5. Another part of this study will focus on how learners perceive and understand *authentic* dictionary definitions written with and without a DV. This study will be performed on definitions in both dictionaries for learners and for native speakers.

RQ6. The final question focuses on factors relevant to comprehension of definitions, such as the approach to DV (restricted vs. unrestricted), the vocabulary burden of definitions, the content of the most frequent and widely used words in definitions, definition length and type. To what extent do these factors affect learners' comprehension of definitions?

3. Studies presented in the book

In order to answer the research questions, I combined a qualitative approach to historical dictionary research with statistical and quantitative methods. I performed a series of studies on a selection of English dictionaries published between 1984 and 2015. All the dictionaries were general-purpose reference works intended for learners or native speakers.

Research questions RQ1 to RQ3 are dealt with in Chapter 4. One of the major studies presented in this chapter compares the dictionaries in terms of the vocabulary burden of their definitions. I use the concept of receptive vocabulary load (henceforth RVL), that is, the lexical knowledge required for unassisted comprehension of written texts (Nation 2011, 136; Nation 2006). Analysis of RVL enables the researcher to estimate the likelihood of definitions being understood by learners. The analysis is intended to reveal potential differences in the difficulty of definition vocabularies between various dictionaries, and to demonstrate whether definitions written with a restricted and unrestricted defining vocabulary differ in this regard. This analysis draws on research examining the effect of text coverage on comprehension (Laufer 1989; Carver 1994; Nation 2006; Webb and Nation 2008; Staehr 2009). Text coverage is the percentage of known word tokens in a text. It is measured on the basis of lexical properties of words, such as frequency, range, and semantic transparency. Frequency is important, but words can occur frequently in a specialist text but never in other texts. Therefore, we also use range as a parameter indicating the distribution of words across various discourse types. These parameters are the main criteria of controlling vocabularies for pedagogical purposes (Thorndike 1921; Horn 1926; West 1953; Xue and Nation 1984; Nation and Hwang 1995; Nation and Waring 1997; Nation and Webb 2011).

The study of RVL is performed using synchronic and diachronic perspectives. The former is adopted to see how different dictionaries differ with respect to RVL, while the latter, how this parameter changed over years in several editions of the same dictionary. The diachronic perspective will be useful in addressing the question of improvements in controlling a definition vocabulary in dictionaries over the time (RQ3). This question will be explored in detail in Chapter 5, which traces how the dictionaries followed their policies of defining vocabulary and why they were not always implemented as stated in the prefaces.

The question of whether there is a core defining vocabulary (RQ2) will be dealt with in Sections 4.3.3. and 4.3.4. The sections compare several DV lists from the learner's dictionaries and focus on overlapping and non-overlapping vocabulary items. A result of this analysis is the Core Defining Vocabulary (CDV). This is a set of 1,520 words that the learner should know in order to understand most definitions in contemporary EFL dictionaries.

One of the major aims of this book is to emphasize that vocabulary control is *indispensable* for the learner's understanding of dictionary definitions (RQ4). However, a study of the effect of a DV on learners' comprehension of definitions is not without difficulties since there are other factors that may interfere with the results. These include definition format, definition style, the complexity of the entry in which a definition is found, and use of abbreviations and other lexicographic conventions. The present research is designed to curb, at least partially, these problems. To that end, I carried out a study on pairs of invented rather than authentic definitions, using the same syntax but a different defining vocabulary. The study tested the effectiveness of these definitions in a comprehension task, showing advantages of keeping the vocabulary under control. It is one of the user-oriented studies which are presented in Chapter 6.

This chapter also reports on an experiment with authentic definitions written with and without a DV (RQ5 and RQ6). The experiment was designed to check the effectiveness of the definitions in comprehension and perception tasks. The subjects, 215 undergraduate students of English Philology, took a multiple-choice test with definitions as aids. The chapter provides an analysis of how the effectiveness of definitions depends on such factors as the approach to defining vocabulary (controlled vs. uncontrolled), use of highly frequent and widely-used defining words, definition format (analytical vs. synonym), and definition length.

To achieve a more complete picture of how dictionaries control their definition vocabularies, Chapter 1 presents early applications of word counts and early research on restricted vocabularies. One of the main themes of this chapter is the pioneering research into the vocabulary control movement conducted by leading figures in this field in the 1920s and 1930s. The chapter demonstrates the outcome of their work, the General Service List (GSL). A link between the vocabulary control movement and lexicography is introduced by explaining the design of West's defining vocabulary used in *NMED*, the first dictionary for EFL learners.

Chapter 2 provides a survey of major projects for pedagogical and professional purposes launched after the GSL, and draws attention to the qualities of the GSL in a comparative perspective. This chapter explores the concept of core vocabulary and presents the relevant linguistic research in this field. The link between vocabulary control and lexicography is further explored by considering the theoretical foundations and practical applications of Wierzbicka's Natural Semantic Metalanguage (NSM).

Chapter 3 traces the tradition of defining in plain English and offers a preliminary account of the practices of controlling a definition vocabulary in early learners' dictionaries. In this chapter, the results of previous studies on vocabulary control are synthesized and research hypotheses are suggested. The research

questions presented earlier in the Introduction are elaborated in the light of previous research.

The book closes with a section on general conclusions, where I summarize the findings and formulate implications for definition writing and DV design.

CHAPTER 1

Pioneering research on vocabulary control

The rise of the early learners' dictionaries, such as the *New Method English Dictionary* (NMED) (1935) and the *Idiomatic and Syntactic English Dictionary* (ISED) (1942), owes a great deal to the so-called vocabulary control movement of the 1920s and the 1930s (Cowie 1998a, 6). The intellectual pursuits of the period focused on how “to reduce the effort required to learn a foreign language by identifying those (relatively few) words which carried the main burden of communication in most everyday encounters” (Cowie 1999a, 15). The underlying assumption was that acquiring a foreign language is greatly enhanced by eliminating the burden of learning too many different words (Coady 1997). The question of how the vocabulary can be reduced to a necessary minimum remained open, and three differing approaches emerged in the course of that period. The first was called objective, the second, subjective, and the third, logical (McArthur 1998a). These approaches will be presented below, along with the General Service List (GSL), the outcome of the collaborative effort to reach a compromise. The chapter begins with an overview of the relevant intellectual activities that had taken place prior to the above period.

1.1 Early interests in vocabulary control

The interest in vocabulary control was not only a 20th century phenomenon. Restricting the number of words to which beginning learners are exposed and then gradually expanding the vocabulary at more advanced stages had been known in language pedagogy for centuries (Cowie 1999a, 16). The principle of gradation advocated by Jan Comenius, a 17th century philosopher and teacher of Latin, assumed a gradual progression from the known to the unknown (Howatt 1984, 49–50). According to DeRocher et al. (1973, 1), as far back as ancient Alexandria, scholars discriminated between frequent and rare words to facilitate the learning of literary Greek by local students. In the 10th century of our era, Talmudist scribes developed frequency counts of the words and ideas found in their manuscripts in order to distinguish between usual and unusual words (Klare 1968, 14). The counts were made for teaching purposes: “to divide weekly Torah

readings into comparable units of comprehensibility” (Zakaluk and Samuels 1996, 41; Lorge 1944).

However, vocabulary control was not limited to teaching. A particularly fruitful area of application of restricted vocabularies was stenography. In 1588 Timothy Bright published a treatise entitled *Characterie, An Arte of Shorte, Swifte, and Secrete Writing by Character*. In order to facilitate note-taking, he devised a system of symbols which was supposed to represent a minimum of 559 words. The symbols were to be modified according to certain principles to indicate similar or opposite meanings (Fries and Traver 1963, 2). The author provided a table of approximately 6,000 words whose meanings were rendered by the words from the minimum list, for example: Agree – *Consent*, Almanake – *Booke*, Antedate – *Before*, and Disagree – *Consent* (Fries and Traver, 2–3). As Bright maintains, he attempted to revive the ancient practice of shorthand to ensure efficient communication in commerce and politics (Kitzes 2006). Bright’s work is considered as the first attempt in modern times at developing a controlled vocabulary capable of expressing all necessary concepts in language (DeRocher et al. 1973, 2).

In the November 1843 issue of the *Phonotypic Journal*, Isaac Pitman, an educational reformer demonstrated how the task of taking notes in shorthand can be made easier by collecting a limited number of words. The frequency counts provided by Pitman were based on a modest selection of 10,000 words drawn from 20 books (McArthur 1998a, 51). The selection included: *saw* (2), *say* (13), *see/sea* (24), *shall* (34), *short* (2), *so* (25), *spirit(ual)* (13) (McArthur, 51). As can be seen, Pitman conflated homophones (*see/sea*) and words related by morphological derivation (*spirit/spiritual*), and made no distinction of whether *saw* denoted a cutting tool or the past simple form of *see*. Although the system was devoid of theoretical grounds, its practical value was appreciated by critics. In a letter to the journal editor, James Biden pointed out that a very similar list had been compiled around twenty years earlier by John Freeman (Biden 1843, 170). The pamphlet in which the list was found was called “A method of teaching adult persons to read; which is designed to obviate their objections and accelerate their progress.” (Biden, 170)

A much more extensive project was conducted in German by F.W. Kaeding (1898), who compiled a list for the purpose of training stenographers. In his *Häufigkeitwörterbuch der Deutschen Sprache*, Kaeding assembled an alphabetical list of word types on the basis of an astonishingly huge corpus of eleven million running words extracted from a variety of sources (McArthur 1998a; Howatt 1984, 257). The word types were accompanied by their frequency counts. Another undertaking to aid stenographers was George Dewey’s *The Relative Frequency of English Speech Sounds*, published in the United States in 1923 (Dewey 1923; McArthur 1998a). In an attempt to establish “the commonest syllables of English”, Dewey

compiled a list of 100,000 words from various types of text, including novels and newspapers.

At the turn of the 20th century, there was a growing interest in undertaking word counts for general educational purposes. An impetus to such activities was British workers' dissatisfaction with the quality of spelling books available at the time and the keen interest to replace them with more objective word counts (McArthur 1998a, 52). The idea was that such counts would aid their pupils in reading and spelling. Another contribution to the field is J. Knowles' *London Point System of Reading for the Blind*, published in 1904 (Knowles 1904). Knowles selected 353 most common words out of a list of 100,000 words found in the Bible and other texts in order to help blind students in reading (McArthur 1998a). Worth mentioning is also R.C. Eldridge's *Six Thousand Common English Words* (Eldridge 1911). In the note to the reader, the compiler explains that the "ultimate aim of the work is the introduction of a limited vocabulary for universal use" (Eldridge 1911, 3). Eldridge believes that his selection of words is "already in frequent use in some one or more of the principal languages and nations of the world including the English" (Eldridge, 3). The vocabulary was supposed to enable people to understand each other "though they had no other words in common" (Eldridge, 3). Apparently, Eldridge was inspired by the old idea of a possibility of developing a universal language that would solve cross-linguistic barriers in communication (McArthur 1998a, 53). The words selected were supposed to occur "the most frequently in ordinary use", and were retrieved from a few different newspapers published in Buffalo on different dates (Eldridge 1911, 3). The words were listed together with the figures denoting the raw frequencies with which they appeared in a given newspaper. All in all, language learning, stenography, and the quest for a universal vocabulary were the main areas of early application of vocabulary control (cf. McArthur 1998).

The 1920s and 1930s were a period of extensive research into vocabulary control. Although it emerged as a promising area of language learning research, it was accompanied by a great deal of controversy. A moot point was the principles according to which a controlled vocabulary should be compiled. One of the approaches gave priority to quantitative parameters of vocabulary selection. Because the parameters were measurable, the approach was called objective. This is what we will be concerned in the following section.

1.2 The objective approach to vocabulary selection

The adherents of the objective approach to vocabulary selection compiled restricted vocabularies according to quantitative criteria of frequency and distribution of

words in a corpus. This approach underpins the works of American researchers, most notably Edward Thorndike and Ernest Horn, whose vocabulary lists were the most extensive in the early 1930s (Gilner 2011).

1.2.1 Edward L. Thorndike

In 1921 Edward Thorndike published *The Teacher's Word Book*, in which he listed alphabetically 10,000 words that young native learners of English were likely to find in their general reading. The book was intended to assist elementary and secondary school teachers in “estimating the commonness and importance of words” (Thorndike 1921, iv), and to guide them in selecting suitable texts for their learners (Battenburg 1994). The words were selected manually on the basis of two statistical criteria: word frequency in the corpus of 4 million words, and range, that is, the number of texts (out of 41) in which a word was found. To express the two parameters, Thorndike devised a single measure, the so-called credit-number. Each credit-number that accompanied a word in the list answered the question of how often the word was used in the corpus and how widely it was used across the texts (Thorndike 1921, iii). Thorndike's corpus contained 41 texts, which made up approximately 4 million words.

From the methodological point of view, Thorndike's decision to take into account range (i.e. the occurrence of a word across the texts) deserves credit, as it reduces the overrepresentation of terms that are technical or genre-specific. They occur relatively frequently in specialized texts but rarely in general ones. This fact has a direct bearing on FL learning: it is more important for learners to learn the words that occur ten times across various books than twenty times in one book. Thorndike's approach addresses the problem of what is representative of the general language. He was well aware of the fact that it is not enough to determine how common a word is, but also how widespread it is in the language. For a word to be representative of the general language, it needs to occur frequently enough across the various sublanguages or subcategories of the corpus.¹

In 1931 Thorndike incorporated additional material to the list, which came from over twenty new sources. Consequently, the list doubled in size, reaching 20,000 words. The book was published in 1932 under the title *The Teacher's Word*

1. Range is a distributional requirement used by contemporary researchers as an important measure for identification of stylistic features of a register. In order to determine so-called stylistic discriminators of a register, scholars determine those lexical bundles that are sufficiently widespread across the texts from that register (Biber et al. 1999; Biber et al. 2003; Biber et al. 2004; Biber 2006; Grabowski 2015).

Book of 20,000 Words. However, as a list to be used in elementary education, it was criticized for including the words of dubious utility for children. As West observes (1935, 9), *escutcheon*, *evinced*, *festal*, and *flaunt* were on the list, but other more useful words were not: *aeroplane*, *marmalade*, *padlock*, *radio*, *slang*.

With the cooperation of Irving Lorge, Thorndike expanded the list further in 1944. The list was published as *The Teacher's Word Book of 30,000 Words* (Thorndike and Lorge 1944). In spite of the shortcomings mentioned above, the book achieved considerable success, and became an inspiration for subsequent language educators until well into the 1970s (see Section 2.1.3). It was treated as an objective reference point for the evaluation of course vocabularies and the construction of vocabulary tests (McArthur 1998a).

1.2.2 Ernest Horn

Ernest Horn was a professor at the University of Iowa in the first half of the 1920s (Schul and Hamot 2011). His interests ranged from the identification of word frequency and the design of school curricula to the compilation of spelling vocabularies (Gilner 2011). In his book *A Basic Writing Vocabulary* published in 1926, Horn identified 10,000 most frequent words in a collection of texts from business and personal correspondences, meeting reports, newspapers, and magazines. His corpus consisted of approximately 5 million running words. Horn was alert to the problem of the “lack of uniformity among the various investigations in the method of tabulating words” (Horn 1926, 17; Gilner 2011, 67). Following Thorndike, he took into consideration both the frequency and the distribution of words across the corpus sample. Because the list was intended to be used for teaching spelling, he deliberately excluded short words of up to three letters, which he found easy to spell (DeRocher et al. 1973). In addition, he left out a few dozens of common words ranked within the 100 most frequent English words.

1.2.3 Lawrence Faucett and Itsu Maki

Lawrence Faucett was the author of numerous EFL coursebooks, readers, teacher's guides, and teaching materials. In 1927 he published a successful book titled *The Teaching of English in the Far East*, which brought him to prominence in the field (Smith 2003, 2015). He was the first to teach a year-long training course for EFL teachers in the UK. During his stay in Japan, he collaborated with Itsu Maki, a researcher and teacher of English.

Lawrence Faucett and Itsu Maki combined Thorndike's and Horn's vocabularies into one list by selecting only those lexemes that are common to both lists. The compilers assigned new values to the words according to “frequency-range importance”. The values varied from 1 to 120, the former indicating words of high

frequency and wide range, and therefore most useful for learning and teaching purposes, whereas the latter, words of minimum frequency-range. The vocabulary was arranged into several classes comprising indispensable, essential, useful, and special words. The compilers believed that their list would become “the essential minimum vocabulary for teaching and testing”. (Faucett and Maki 1932, ii). In 1932, they published it under the title *A Study of English-Word Values Statistically Determined from the Latest Extensive Word Counts*. The Faucett-Maki-Thorndike-Horn list would later serve as a starting point for an Interim Report list (see 1.5.2) (Faucett et al. 1936).

1.2.4 The strengths and limitations of statistical word-counts

The studies of distribution of words yield support to the importance of frequent words for language education. According to Zipf (1935), few words are very frequent in the language, whereas most words are relatively rare. The former include function words such as *the*, *a*, *of*, *and*, and *at*. Because they are often repeated, they account for a very high proportion of the entire vocabulary. Research into the distribution of words in the Brown corpus illustrates this point. In the frequency-ranked list of word types from this corpus, the first 10 words constitute 23% of the total count of tokens (Baroni 2006), the top 100 words, 45% of the corpus, whereas the top 2000 words, as much as 80% of it (Bongers 1947; Francis and Kučera 1982). According to Gilner (2011), regardless of the source from which a text originates, around 2,000 words constitute 70% to 95% of all running words. In other words, the reader has a high chance of encountering the same words throughout a text. These words include not only function ones, which are the most frequent according to Zipf, but also numerous lexical words that are widely used in the language. This finding provides a sound basis for using frequency as a major criterion for vocabulary selection in language learning. However, it will be shown below that the results of such a selection can sometimes be at variance with the teacher’s judgement as to what counts as useful for the learner.

Although the question of vocabulary selection was seemingly resolved with the statistical criteria, the methods used by Thorndike and his followers came under severe criticism. Critics highlighted several points, such as serious weaknesses of the source corpora, problems with determining the unit of counting (i.e. what counts as a word), and a mismatch between the results of the statistical method and teachers’ intuition.

As McArthur (1998a, 53) observes, the reliability of Thorndike’s counts is diminished by the limitation of the corpus used, both in terms of its size and text selection. The 41 books provided over 4 million words, of which three-fourths (3 million words) were from the Bible and English classics, and the remaining

one-fourth came from letters, books for children, elementary school textbooks, daily newspapers, and general reading (e.g., books on cooking, farming, the trades, and sewing) (Thorndike 1921, iii). The obvious weakness of this corpus, which Thorndike collected with elementary and secondary school teachers in mind, was that it was selective and unrepresentative of the English language as a whole. With a strong bias towards written literature, the texts were selected as, what the author perhaps considered to be, a standard for English education (McArthur 1998a, 53–5).

Other problems resulted from the difficulties in establishing what counts as a word. The need for a consistent treatment of linguistic signs was especially acute in the course of expanding the list by adding materials from different sources. The compiler counted homographs as if they had been the same words, in the fashion similar to that of Pitman, who in his early work of 1843 had failed to differentiate between homophones (McArthur 1998a, 53). Furthermore, Thorndike ignored the non-uniform treatment of word-forms in the source lists; for example, *am*, *is*, *are*, *were*, and *was* were treated either as separate or the same words, depending on which list they appeared in (McArthur 1998a). As McArthur points out (1998a, 54), the conflation of the lists contributed not only to the increase in the size of the corpus but also to the error affecting reliability of counts, resulting from the lack of a theory of language. The problem of how to determine various lexical categories would be the main strand of enquiry undertaken by Palmer (1929a, 1929b; see Section 1.3.1).

Another serious flaw of the objective approach was that the results did not always tally with the teacher's intuition (Richards 1974). Firstly, the approach yielded different results, depending on the corpus used. While the first thousand words were determined "with small likelihood of error", the list beyond that boundary would be significantly different if the sources had comprised scientific and technical texts (Cowie 1999a, 16). Secondly, as Richards (1974, 72) observes, the method does not do justice to the common concrete words which are often used in the classroom discourse. For example, *soap*, *soup*, *basin*, *bath*, *oven*, *dish*, *chalk*, *blackboard*, *pencil*, *umbrella*, *camera*, *stomach*, and *trousers* are among the words omitted from the first 2,000 items of Thorndike and Lorge's list, which in fact was compiled with the needs of children in mind. This is surprising, as many of these words represent objects that children normally encounter in the classroom.

It is worth mentioning in passing that hand, chalk, blackboard, and other objects that are normally found in the school environment are rarely encountered by adults in everyday life. This is exactly why the compilers of the GSL (see 1.5.2) excluded words for these objects from the list because the list was intended to be of general service to foreign learners of all ages. As they explain, the words "might, without detriment, be forgotten as soon as the child leaves school." (Faucett et al. 1936: 15).

In turn, as Richards (1974, 72) continues, within the first 1,000 words of frequency-based lists compiled by Thorndike and others (i.e. Horn and Dewey), one finds abstract verbs of doubtful utility for foreign learners: *construct, commit, charge, approve, dispose, err, execute, stock, lack, and grant*. The words missing from that range are *soft, hungry, animal, catch, corner, dirty, and worry*. The above inconsistencies led critics to argue that word frequency alone is insufficient for vocabulary selection, and that there is no straightforward relationship between word frequency and its usefulness for a learner (Richards, 72).

One implication of the statistical studies of distribution of words is that by learning the top 2,000 words, learners can understand the vast majority of the lexical content of any text. However, as Richards (1974) argues, this is not tantamount to the learner's full understanding of a text. He goes to explain that although learners who are familiar with these words (i.e. the top 2,000) are able to recognize 80% of the text vocabulary, they may fail to understand the text, as the information essential to comprehension may be hidden in the remaining 20% of the vocabulary. It is in these outsiders that the essential information may lie. This type of vocabulary tends to be high in information content (see Lyons 1969), and thus may be crucial to communication. Richards (1974, 72) illustrates this point by comparing the following two sentences:

- (1) *He came in and put a thing on the bed.*
- (2) *Mr. Smith shuffled into the bedroom and placed a jacket on the bunk.*

The first sentence uses the words of high frequency, that are of general utility to learners. They express the most general meaning, and are "as culture-free as calculus, with no literary, aesthetic or emotional aspirations" (Quirk 1982, 43). Nevertheless, as Richards (1974) argues, it is the second sentence that carries more information. This is due to the use of semantically richer words. Because the low-frequency words, *shuffle, jacket, and bunk*, are hyponyms of the high-frequency words from the first sentence (i.e. *come, thing, bed* respectively), the second sentence contributes more information to the communication than the first sentence. The researcher concludes that foreign language learners need not only the words that are highly frequent and widely distributed in the language, but also other words that cannot be identified by frequency and range. There were other voices that called the objective approach into question. The voices were given by the followers of the subjective approach.

1.3 The subjective approach

While the American interest in vocabulary control was essentially centred on statistical word-counts, British researchers relied more on their pedagogic intuition

and experience. They believed that experts' judgments provide more satisfactory grounds for compiling controlled vocabularies than any frequency-based decisions (Cowie 1999a). The leading figures of this approach were Harold Palmer and Michael West.

1.3.1 Harold Palmer

A great deal of initial work within the framework of the subjective approach was done by Harold Palmer, a language researcher and educator. Palmer taught English in Belgium, then in London, and finally in Tokyo, where he served as director of the Institute for Research in English Teaching (IRET). He took a dim view of "objective" word counts offered by Eldridge and Thorndike, arguing that any statistical work needs "precise definitions" of the items under investigation (Bongers 1947, 81). Palmer did not accept the fact that existing word counts completely ignored homonyms, multiple meanings of words, and relationships between derivatives and their roots (Cowie 1999a). He believed that none of these factors can be overlooked when a vocabulary is to satisfy the learner's receptive and productive needs. It is remarkable that the factors are still relevant to the compilation of contemporary vocabularies for teaching purposes. Some of the factors are by no means easy to control, in particular the multiplicity of word meanings remains a cause for concern for compilers of dictionaries and teaching materials (Meer 2004; Lew 2013). The disambiguation problem is caused by the inherent semantic ambiguity of words. While many words have meanings that are relatively fixed, others acquire a large part of their meanings from context (Lew 2013, 286).

As a language teacher and advisor to the Japanese Ministry of Education, Palmer engaged in developing graded word-lists which served as a lexical base for writing simplified texts for different-level groups of students. The words lists were prepared in such a way that the students could learn English key words in stages, from around 300 to 3,000 most important words. This strategy ensured a limitation on the number and scope of vocabulary items to which learners were exposed (Battenburg 1994, 136). In the course of revisions, the 1,000-word minimum list appeared to be the most popular. Palmer believed it to be "the necessary minimum", which "would account for the great bulk of any text composed in 'everyday English'" (Cowie 1999a, 16). According to his estimates, the words on his 1,000-word list would cover 85% of a text, while 3,000 words would account for 95 per cent of a text. The latter estimate was proved accurate by Bongers (1947) in the study of Shaw's *Doctor's Dilemma* (96.1%), Arnold Bennett's *The Card* (95.2%), and a number of other literary works (Howatt 1984, 257). Palmer maintained that candidates for key words should be established by experienced teachers themselves using, what he called, "vocabulary sense" (McArthur 1998a, 58). This opinion was shared by Michael West, a fellow researcher working in Bengal (58).

In 1927, when Palmer accepted a proposal from IRET to compile a controlled vocabulary for Japanese learners of English, he set out to establish various lexical categories (Cowie 1999b, 4). As Cowie (4) points out, it is Palmer's "almost obsessive preoccupation with the definition of lexical units" that shaped his vocabulary as well the future learner's dictionary. The categorization problem was addressed in his two 1929 papers, entitled "What shall we call a word?" and "What is an idiom?" Palmer developed a typology of form-classes: *miologs*, which referred to what we would now call morphemes; *monologs*, a name for contemporary word forms; and *polylogs*, a name for collocations and phrases (Howatt 1984). In his quest for lexical categories, the researcher attempted to solve a perennial problem in lexicography, namely which vocabulary items merit headword status and which do not (see McArthur 1998a, 58). He coined the terms *headword* and *subword* to make a distinction which was relevant not only to the conception of his word-lists, but also to FL dictionaries. This conceptual distinction had enormous influence on the selection of the vocabulary for the GSL (Gilner 2011). *Headword* represented the main entry, while *subword*, the items associated with the headword, in particular derivatives, compounds, and idiomatic expressions (McArthur 1998a, 58). An example of Palmer's approach can be the following entry:

accept: v
 acceptance: n
 acceptable: adj
 unacceptable: adj
 (Bongers 1947)

As we can see above, Palmer grouped words together in order to show that they are linked by derivational affinities to the common core. His approach to the entry structure foreshadowed the type of microstructure that would become a long-time standard in *OALD*. The nested structure was used in *ISED/ALD1* and its numerous revisions up to the end of the 20th century.^{2,3} Palmer was the first to raise the issue of the lexicographic presentation of words for EFL learners. His lists were

2. The nested type of microstructure was used in the early EFL dictionaries, but as such, it was not novel to popular dictionaries for native speakers. For example, concise dictionaries at the turn of the 20th century, such as *Chambers's Twentieth Century Dictionary* (1901) and *Concise Oxford Dictionary* (1911), had placed derivatives, compounds, and multi-word expressions under a common core (root or base word). However, this layout resulted from the policy of space saving rather than the need to make the user aware of derivational affinities.

3. Until the 5th edition of *Oxford Advanced Learner's Dictionary* (*OALD5*1995), the dictionary favored a nested type of microstructure, with derivatives and compounds being clustered together under a common core.

designed not merely to establish which words were necessary for communication, but chiefly to reduce the burden on the learner's memory. The latter was to be achieved by clustering words around their cores. Certainly, this strategy was productively useful, as it made the user aware of derivational relationships among words. To some extent, it was also useful for decoding, as the awareness of such affinities would help the reader understand unfamiliar but predictable derivatives. Nevertheless, Palmer's layout gave a higher priority to production than reception. This information layout contrasted with that of *LDOCE* (all editions), a reception-oriented dictionary which assigns separate entry status to complex and compound forms as well as phrasal verbs.

Palmer was the first to compile a word-list by collecting word families rather than just spelling forms (Gilner 2011). This methodology stood in contrast to that of Thorndike, whose list of 1921 had been an alphabetically arranged set of spelling forms selected on the basis of their occurrence in the corpus. Palmer's methodology would be followed by West (1953), as well as more recent researchers including Xue and Nation (1984) and Coxhead (2000) (see Chapter 2).

Palmer collaborated with A. S. Hornby in a research project which led to the compilation of 1,000 word-list published as *Thousand-Word English* (1937). The list provided the elementary words that foreign learners of English "would do well to learn first" (Palmer and Hornby 1937, 7). The vocabulary was regarded as "the most useful and practicable of its kind for writing or adapting reading materials" (Cowie 1999a, 19). The vocabulary selection, however, was not based purely on the compilers' intuitions. As they admitted, the method was "partly subjective, partly objective, and partly the empirical (or experimental) method" (Palmer and Hornby 1937, 12). The compilation process began with a draft list proposed by Hornby "for the purpose of re-writing stories of more ambitious content" (Palmer 1936, 21; cited in Cowie 1999a, 19). The list was verified subjectively on the basis of the compilers' experiences as teachers and writers of teaching materials (Stein 2002b, 5). Then the researchers made sure that the list included all the common words found in the objective lists available at the time, in particular Thorndike's *Teacher's Word Book* and Palmer's 3,000-word list that was incorporated into the *Second Interim Report on Vocabulary Selection* (Stein 2002b, 5). This part of the selection was carried out following the objective criterion. Finally, they tested the vocabulary of the list by rewriting, or paraphrasing, various texts within this vocabulary. Meanwhile, they kept revising the list in order to make sure that the words selected were absolutely useful (Cowie 1999a). Paraphrasing was the fundamental method for establishing the most useful words by West (see 1.3.2), as well as Ogden and Richards (see 1.4).

Palmer and Hornby's *Thousand-Word English* was not merely a list of alphabetically arranged words, but a compilation of elaborate entries containing a range

of information types, from inflectional forms and derivatives to meanings and phonetic transcriptions, all gathered together under a common base or root word. An example entry is shown below.

APPEAR [ə'piə], *v.*
 (1. = come into sight)
 (2. = seem)
 appearance [ə'piərəns], *n.*
 disappear [disə'piər], *v.*
 disappearance [disə'piərəns], *n.*
 (*Thousand-Word English* 1937)

A similar approach was adopted by Palmer and his collaborators in compiling *A General Service List* (GSL) in 1936 (see 1.5.2). Published as part of the *Interim Report on Vocabulary Selection* (*Interim Report*), the GSL also offered illustrative examples and collocations, the latter selected by Hornby. Such “structured lexicons” served as a framework for the compilation of learners’ dictionaries (Cowie 2009, 389). Although neither of the lists provided syntactic patterns, they had a great deal of encoding potential, of which Palmer was not fully aware at that time (Cowie 1999b, 6).

In 1938, Palmer compiled *A Grammar of English Words* (GEW). The book was essentially a grammar designed specifically to meet encoding needs of learners, though, as the editor remarks, it could also be entitled *A Grammatical Dictionary of English Words* (Palmer 1938). It was a collection of only 1,000 core words that for various reasons presented considerable difficulties for foreign learners. Being a small volume of 300 pages, it provided a great deal of productively-useful information, which ranged from illustrative examples, verb patterns, glosses specifying the context of use, and alternative words embedded in supporting examples, to part-of-speech (POS) labels. For many words the compiler deliberately omitted definitions on the assumption that the meanings were sufficiently clear from illustrative examples. The definitions given served as brief indications of meaning, in line with the purpose of the dictionary.

It is worth noting that Palmer’s system of construction patterns used in *GEW* was a result of several years of his research in this area. At the early stage of the research, he observed that “it is not so much the *words* of English nor the *grammar* of English that make English difficult, but that that vague and undefined obstacle to progress in the learning of English consists for the most part in the existence of so many odd comings-together-of-words” (Palmer 1933, 13; cited in Cowie 1999b). Palmer paid particular attention to complementation and categorization of verbs according to various usages. In 1932, he designed patterns for his students to guide them on verb syntax, and published them in

Some Notes on Construction-Patterns. A few years later, he greatly elaborated the scheme and used it in *GEW*. The scheme was further refined by Hornby for a much larger dictionary, *ISED* (Rundell 2008a). Hornby improved the presentation of verb patterns by arranging them in tables and dividing them into transitive and intransitive ones (Cowie 1999a, 33). Both Palmer and Hornby were committed to meeting the encoding needs of learners, and *GEW* and *ISED* embodied the results of their work. West set a different goal: his primary interest was in developing learners' receptive skills, and vocabulary control was central to this task.

1.3.2 Michael West

Michael West (1888–1973) is a crucial figure in the history of vocabulary control. He was the principal compiler of *NMED* (1935), the first dictionary for EFL learners prepared with a restricted defining vocabulary. By the time *NMED* was published, West had already gained a reputation as the leading British writer of EFL textbooks (Smith 2003).

While working as a colonial educator in the Indian Education Service, he was posted to Bengal in 1912. The fundamental question that West addressed was whether the vocabulary necessary for reading was different from that of speaking. West regarded reading to be a priority for bilingual learners, so he focused his efforts on making reading material more intelligible to his Bengali students (Cowie 1999a, 22). He experimented with a series of textbooks for FL learners, showing how strict vocabulary control can help students acquire reading skills relatively easily without the presence of a teacher. The results of the experiment were reported in *Bilingualism with Special Reference to Bengal* and in *Learning to Read a Foreign Language*, both published in 1926. He maintained that readability of literary texts can be improved by adhering to two principles. The first principle holds that rare, dated, and specialized words should be replaced with common ones, and the second, that distributing new words at slow rate allows sufficient time for the reader to absorb them (Cowie 1999a, 7; Schmitt 2000).

In 1927 he published the first volume of *The New Method Readers*, a series of simplified texts written within a restricted vocabulary (West 1927b). The readers were graded by stage, from early beginner to advanced. The vocabulary of the first five volumes consisted of 1,779 words. It had been identified through experimentation by constant rewriting of literary texts, substituting common words for rare ones, and keeping a record of the words used (Stein 2002b, 17). These strategies allowed West to make adjustments to the texts being written as well as the vocabulary list. The list was considered to be the adequate vocabulary for general purposes (Stein, 17).

The preparation of a simplified story involved decisions regarding the use of low-frequency words, many of which were, as he noted, unavoidable. To that end, West performed an initial analysis of the lexical content of the story. As he explained, “[t]he story must first be told through mentally in order to observe what words of low frequency are likely to be inextricably involved in the plot” (West 1927a, 23). He looked for the words which he could not do without and accepted them within certain limits, as he pointed out: “if there are too many the story must be rejected” (West 1927a, 23). While writing the first draft, he selected new words from the frequency lists available at the time, including Thorndike’s (see also Stein 2002b). He also referred to the vocabulary of the other books in the series to check, at a regular pace (i.e. “every third or fourth word”), if a word had already been used and if it had been employed in the same meaning (West 1927a, 23). New words were systematically highlighted in boldface and recorded on the margin to draw the reader’s attention. The final draft was read through in order to detect any awkward expressions resulting from the use of low-frequency words (West 1927a).

When West laid down the criteria for the minimum adequate vocabulary of a school course, he was careful not to restrict the learner’s profile, emphasizing that the words should be usable for a wide range of learners:

The vocabulary should be usable by a boy or girl in a moderately civilized environment – not a big city nor a mere country village. The principle of Negative Selection was employed: a word or item should not be included unless it was such as to be useful to boys and girls, in town and country. (West 1960, 39)

West’s minimum adequate vocabulary included “common environmental words”, that is, the words which were “a part of the environment of all men”, for example, *sleep, eat, bed, food* (West 1954). These words were independent of students’ age, country, culture, etc., and thus were more predictable than the words related to their interests and preoccupations:

Obviously, it is possible to select a list of Content Words provided that we have only one pupil and know all about his present and his immediate future. ... for example, the boy will not need names of articles of female apparel, but the girl will need names of male clothing because she will have to mend them. Add to our urban boy and girl a country lad: he may need more agricultural words, but not need so many of the luxuries of the city. ... Suppose we have to consider not three pupils, but thirty? –or all the pupils ... in the whole world? ... The larger the area and number of pupils the less one can predict. (West 1954)

West recommended that the minimum vocabulary “should be taught and remembered” by the student, and that the words qualifying for this vocabulary had to be useful for a wide range of learners:

‘Will it be useful to all my pupils after they leave school? Will it be often necessary?’ If so, include it and ensure that it is remembered. If not, treat it as an ad hoc word, something necessary for this lesson but not worth permanent storage. (West 1960, 39–40)

The preparation of the vocabulary for speaking and writing incorporated the same method of paraphrasing. West tested the substitutional power of the vocabulary of the list by rewriting plays, novels, adventure stories, business conversations and other texts that a learner was likely to produce (West 1954, 125). He checked whether the listed words were capable of expressing all the ideas mentioned in the text, noting the words that were “absolutely essential” for the context but missing from the word-list. He also employed this strategy in writing definitions for *NMED*.

1.3.2.1 *West’s defining vocabulary*

The vocabulary of the *New Method Readers* was only a starting point for a more ambitious project. The aim of this project was to “discover what smallest vocabulary is adequate for the writing of a dictionary, – what least number of words is capable of defining or describing the idea of all other (ordinary non-technical) words” (West 1935: 20). Another question that West asked is still pertinent to contemporary lexicography: “If you are writing the explanations of English words in English, how do you know that the foreigner will know the explaining words which you use?” (West 1935, 7).

In the initial draft of the dictionary, West and his collaborator James Endicott attempted to write definitions using the vocabulary of the first five volumes of the *New Method Readers*, that is, 1,779 words (West 1935, 20). This vocabulary served “merely as a starting point” for writing definitions because the compilers wanted to “avoid prejudice in favour of [their] original vocabulary” (20). At the same time, they set a goal of discovering which words merit their existence in the dictionary. As the dictionary progressed, new defining problems came to light, prompting the lexicographers to revise the original list further. As West (1935, 21) explains, the main part of the dictionary was written twice: “the first time on a smaller scale with less idioms, and the second time on a more elaborate plan”. The compilers even experimented with a small set of vocabulary items, consisting of 1,200 words, but this attempt turned out to be “a failure” (West, 21).

In the year of publication of *NMED* (1935), West also wrote *Definition Vocabulary* (West 1935), in which he laid down the selection criteria for his DV. In this account, he raised several problems relevant to learner’s lexicography. He identified shortcomings of definitions in native-speaker dictionaries, such as circularity, defining *obscurum per obscurius* (e.g., **digenite** “a variety of *chalcocite* or *copper-glance*”), and the use of a string of near-synonyms that neither match the headword

in meaning, nor are intelligible to the user (e.g., **seemingly** “fittingly, becomingly”) (West 1935, 8). He noticed that many common words were defined by means of words which the user was unlikely to know. In order to avoid misunderstandings, West suggested the use of words that are assumed to be known by the learner, and ideally each word in one meaning only (West 1935, 8). Therefore, he suggested that editors establish beforehand whether the word *wrong* should be used in the meaning of “moral wrongness” or “incorrectness.” According to West, the criterion of frequency was unsatisfactory for identifying the words familiar to foreign learners. He noticed that the first eight thousand words on Thorndike’s list include *escutcheon*, *evince*, *festal*, and *flaunt*, apparently because they were frequent in specialist texts, but surprisingly there was no *aeroplane*, *marmalade* or *radio* (West 1935, 9). West was well aware of the fact that a restricted defining vocabulary may result in definitions being excessively lengthy. He noted that a similar problem arises in the case of the so-called double definitions, to which lexicographers resort when they define semantically complex words. Definitions of such words require the use of related words of low-frequency (e.g., *pickle* and *vinegar* in the definition of **gherkin**), which in turn need to be defined for the learner. West (1935) considered the inclusion of such words into the DV, concluding that they may extend the word list. On the other hand, he argued that, if these words keep recurring in other definitions, learners “may automatically acquire or deliberately learn [them] in the course of using the dictionary” (1935, 5).

In the *NMED* DV, West and Endicott included “the commonest and reasonably regular prefixes and suffixes”, which could be attached to the base words already on the list (West 1935, 16). In this way, the compilers assumed that learners would be able to infer the meanings of derivatives used in definitions. The user was supposed to know the affixes prior to consulting the dictionary. Obviously, this strategy worked as long as the meaning of a resulting word was fully predictable from its parts, and West realized the consequences of using derivatives and compounds whose meanings were not directly retrievable from the structure (Cowie 1999b, 9). By using transparent items in definitions, a great deal of space could be saved and defining vocabulary could be kept short.

In the course of the dictionary project, West and Endicott kept a separate list of the words that had been excluded from the original 1,779 word-list but which turned out to be desired for some other reasons. The compilers listed the “desired” words along with other words for definitions of which they were needed, sometimes providing brief remarks explaining the reasons. For example, *flavour* was wanted for definitions of *spice* and *herb*, because *taste* was insufficient for that purpose; both *hiccup* and *sneeze* were needed because “[r]eflex actions are difficult to define”; and *cure* was needed for definitions of *antibody*, *hospital*, and *sanatorium*, but, as the remark explains, both *cure* and *heal* were needed because “[h]eal is cheaper; [c]ure is wider” and “one cannot cure a wound” (West 1935, 73).

By the time the dictionary was completed, the lexicographers identified in the original vocabulary 552 “doubtful” words which were “open to objection” (West, 21). They also added 61 words which they “could not get on without, that is, it was impossible to give a satisfactory definition in a large number of cases without bringing in one of these words” (West, 20). In this way, the following words were “forced into” the vocabulary:

acid, active, amuse, behaviour, belief, bitter, brain, card / cards, central, chalk, cheap, clever, continual / continually / continuous, copy, cotton, curve, dig / dug, electric / electrical / electricity, enable, enclose, engine, female, fold, frame, furniture, gas / gaseous, heap, hook, imaginary / imagination, insect, instrument, jelly, lawyer, lend / lent, male, medicine, medical, metal, mixture, muscle, neat, note (of music), noun, ornament, participle, pig, plural, police / policeman, post (e.g., wooden), priest, quality, race (running), rail / railway, relation (family and science), rubber, science / scientist / scientific, second (time), shelf, skill / skillful, slide, snake, solid, sour, steam, stem, surface, tense (of verb), theatre, thread, twist, university, upright, vegetable, weigh, willing, wire.

Some of the above words were considered to be useful as superordinate terms (e.g., *behaviour, belief, engine, furniture, insect, instrument, metal, noun, quality, relation, science, skill, surface, vegetable*) (Cowie 1999b, 10; Smith 2003, xx). The words were useful in defining other words, because they were neither too specific nor too general, and thus were capable of encompassing a great deal of concepts, both concrete and abstract. Cowie (1999a) observes that as many as 17 of the above additions feature in the defining list of *OALD5* (1995) and all, but one (*ornament*), in the list of *LDOCE3* (1995). It is worth adding that some of them, such as *quality*, had already been extremely frequent in native-speaker dictionaries (Atkins and Rundell 2008). Other categories of words needed for defining were the names of grammatical classes (e.g., *noun, participle, tense, plural*).

In *Definition Vocabulary*, West reports on an experiment in which he and his collaborators evaluated words according to the range of reference. The words ranked highest were those that entered into a “number of subjects of conversation or discourse” (West 1935, 22). This is how he compares three words for kinds of fruit:

it is evident that Orange is the more valuable word, since it is a fruit, a colour and a shape, and its flowers have romantic associations. Apple is less valuable as a colour, of no importance as a shape, but is has a slight extra importance in respect of story and fable material. Banana is merely a food, with a slight value as a shape
(West 1935, 23)

West adds that word range may not necessarily correlate with frequency. He finds that while a high-frequency word *butter* is “merely a food” and falls into very few

categories, a low-frequency word *superfluous* “might apply to anything” (West, 23). The above experiment served as evidence for West’s decisions regarding the selection of the DV.

In the final stage of the work, the compilers compared the vocabulary with ten other word lists, including Ogden’s *The Basic Words*, Horn et al.’s *Vocabulary of Children Before Entering the First Grade*, Palmer and Hornby’s *Standard English Vocabulary*, and Rejall’s *Commonly Used English Words on Adult Education*. The lexicographers identified the words which figured on these lists but were missing from their vocabulary. Although none of the words was admitted to the final vocabulary, they selected 237 items as worthy of inclusion, for example, *almost, charge, cheap, false, information, reply, situation, toilet, and trip* (West, 24 and 93).

By adopting the *Readers* vocabulary and constantly revising and refining it, West and Endicott arrived at a relatively small set of words, as judged by modern standards (note that *OALD5* claims to use over 3,400 items). The vocabulary consisted of only 1,490 items. As the editors explain, the vocabulary serves as the metalanguage for the definitions of 23,898 items, so that “Anyone who knows these 1,490 words will be able to understand every explanation given in this dictionary.” (West and Endicott 1956 [1935], iii). In addition to the restrictions on defining words, they also limited the number of words being defined to a subset of the lexicon. The dictionary was intended for intermediate learners as an aid to reading (Cowie 2009). A revised version of *NMED* appeared in 1965. Its title *An International Reader’s Dictionary* (West 1965) explicitly indicates the function of the dictionary as a companion to simplified readers.

1.4 The logical approach

In the 1920s Charles Kay Ogden and Ivor Armstrong Richards were preoccupied with the development of a controlled vocabulary on principles somewhat different from Thorndike’s statistical counts and Palmer’s teaching experience. Ogden and Richards experimented with paraphrasing complex words in order to establish a limited set of words that keep recurring in the paraphrases and that are most useful for defining purposes. In this way, using the logic of substitution, they arrived at a self-contained core vocabulary of English. According to Ogden, language is organized hierarchically, with different levels of words. Words of lower levels, which are of importance only to specialists, can be replaced with those of a higher level (Gordon 1994, vol. 2, 302–303). Hence, “*quadruped* is redundant in a vocabulary which contains *animal, four, and leg*” (302–303).

According to Richards, the idea of restricting a vocabulary for general communicative purposes emerged in the course of writing Chapter 5 of the *Meaning of Meaning* (1923) by Ogden and Richards:

We were comparing definitions – definitions of anything from a table to a force and from a rabbit to a concept – and we were struck by the fact that, whatever you are defining, certain words keep coming into your definitions no matter how diverse the things you are defining. This suggests that there might be a limited set of words in terms of which the meaning of all other words might be stated. If so, then a very limited language is possible, a language which would put a description (using only this limited set of words) in the place of any word outside this limited set. That, of course, is an old idea, the idea of the ‘universal characteristic’ that was intriguing Leibniz, Bishop Wilkins, and others in the century before the Royal Society of Arts was founded. (Gordon 1994, vol. 2, 13–15)

The vocabulary project came to fruition in the late 1920s, and resulted in the development of a controlled language known as *Basic English*. It was a restricted subset of English developed for foreigners learning this language and anybody trying to communicate across national borders. Basic English was to “take as little of the learner’s time as possible” (Ogden 1931, 9). It was intended as a “system in which everything may be said for all the purposes of everyday existence: the common interests of men and women, general talk, news, trade, and science” (Ogden, 9). Basic was an acronym for “British American Scientific International Commercial”, a name that explicitly highlighted both the target readership and the simplicity of the system for foreign learners.

The main principles of this language were laid down in *Basic English: A General Introduction with Rules and Grammar* (1930). Basic made use of an extremely limited vocabulary of 850 words and a simple syntax. The brevity of the system was one of its most noteworthy, though strongly criticized, features. As Ogden argues,

The fact [...] that it is possible to say almost everything we normally desire to say with the 850 words on the frontispiece, which occupy about three-quarters of the space on the back of an ordinary sheet of business notepaper, is not without significance. (Ogden 1930, 7)

The vocabulary was divided into the names of things (i.e. nouns), names of qualities (i.e. adjectives), and operation words (i.e. mainly function words and verbs). The names of things constituted the largest part of the vocabulary: 400 general words, such as *account*, *act*, *addition*, *adjustment*, and 200 picturable words, such as *angle*, *ant*, *apple*, *arch*. There were 150 names of qualities, and 100 operation words. The latter included a number of function words, such as *about*, *across*, *after*, *against*, *as*, *for*, *of*, *a*, *the*, *I*, *here*, and only 18 verbs, predominantly lexical ones:

come, get, give, go, keep, let, make, put, seem, take, be, do, have, say, see, send, may, will. It is the verbs which vividly demonstrate Ogden's reductionist approach to the selection of the vocabulary. As John Russo notes,

Ogden finally needed just sixteen [lexical] verbs, [...] plus twenty prepositions, to take over the work of innumerable verbs. Abandon, abdicate, abjure, cede, desert, desist, forego, forsake, ... relinquish, renounce, resign, vacate, withdraw and yield could be removed; give up could take their place".

(Russo 1989, 400; cited in Koenke 2004, 93)

The verbs combined with prepositions or adverbials were considered very useful for the learner, for they conveyed the meanings of numerous verbs outside Ogden's list. In the words of Richards, "[i]f we jump, we go *up* and *down*, if we raise a building, we put it *up*, and if we raze it, we take it *down*. If we donate a thing we give it *to* someone or something" (cited in Koenke 2004, 93 and Podhajecka 2016, 374). Ogden's reductionist approach to verbs has been praised by Fries and Traver (1963, 90), who maintain that restricting the number of verbs to a necessary minimum is "not only theoretically sound but helpful from a practical point of view." As they argue, verbs display a large number of meanings imposed by context, so their exclusion makes sense because it reduces significantly the learning burden of the list. Ogden admitted phrasal verbs, metaphorical extensions of verbs, their idiomatic usages, and semantically opaque compounds. All of them were the targets of criticism.

Ogden did not specify meanings in which the words were used, though simplicity and clarity required that the words be used only in fundamental meanings (Chruszczewski 2006, 18; Crabbe 2017). Neither did he restrict the range of possible derivatives and word combinations. However, Ogden allowed for the use of a limited set of affixes: plural *-s*, negative *un-* for adjectives, present and past participles, adverbial *-ly*, and agent *-er*. Thus, with the aid of a few word-formation rules, the learner could create new words from those already on the list; for example, "*designer, designing, and designed, from design, or air-plane from air and plane*" (Ogden 1931, 9). The compounds of the following structures were allowed for: verb plus particle (phrasal verbs), noun plus noun, and adjective plus noun. Syntax was simplified by using active sentences formed on the following pattern: "I will put the record on the machine now" (Ogden 1931, 10). This sentence structure was supposed to ensure that words were used in their "right and natural" places (Ogden, 10). Ogden maintained that Basic "could be learnt in a week or 'at worst' in a month, and once acquired, could be used to express any meaning that could also be expressed in normal English" (Howatt 1984, 251).

As mentioned earlier, Ogden arrived at the basic vocabulary using the method of substitution rather than drawing on any statistical data or teaching experience.

His method was based on paraphrasing a word within other basic words or replacing it by a word “at a higher level” (Gordon 1994, vol. 2, 302–303). Therefore, it was acceptable to dispense with certain words by using “a clear but somewhat clumsy substitute” (Gordon, 269). The inclusion of a pair of opposite words in the vocabulary would not make sense because one of them can be explained in terms of the other. Thus, the following words were excluded from the vocabulary:

- all words (with the exception of common objects) which can be defined in not more than ten other words,
- words which, whether capable of definition within the set limit or not, are primarily emotive rather than referential, according to the distinction established in the theory of symbolism,⁴
- words which, though not capable of definition within the set limit, have chiefly a literary or stylistic value,
- words which, whether capable of such definition or not, are used in contexts too abstruse for the level of general communication. (Gordon, 273)

Interestingly, Basic English was “remarkably similar” to the GSL (McArthur 1978, 60). This is perhaps because the lists shared a large proportion of highly frequent and widely used lexical items.

Nevertheless, Basic English provoked considerable controversy in the 1930s and the early 1940s (Howatt 1984, 250). In 1934, a harsh criticism came from West, Swenson, and others who published a paper under the title “A Critical Examination of Basic English”. The researchers disapproved the system for using considerably more than 850 words and for employing the words in a great variety of meanings. According to West and his collaborators, the alleged number of words was “a considerable understatement” (West et al. 1934, 10), as the actual number reached 1061, and included 76 international words; 48 names of days, months, and numbers; 18 pronouns, and 18 irregular forms of verbs. The vocabulary, said West, failed to cater for the needs of all groups of learners. Arguing against the alleged usefulness of Basic for learners of all ages, West pointed out that numerous international words that Ogden permitted, such as *chauffeur*, *dynamite*, *radium*, *torpedo*, and *volt*, can hardly be known to non-European children (West et al. 1934, 10). The critical voices were also supported publicly by Palmer (Howatt

4. In *The Meaning of Meaning* (1923), Ogden and Richards introduced their famous triangle of meaning, distinguishing between symbols, thoughts, and things. This distinction underpins their notion of symbolism. They argued that words do not communicate anything by themselves, but stand for something only when used by the thinker. They drew a dividing line between the referential use of words and their emotive function (Nerlich 1992).

1984). Recent critics maintain that Basic produces unnatural English. For example, Carter (2012, 41) points out that the system is not suited to spoken interaction because of a lack of the words that typically occur in social communication, such as *good-bye* and *thank you*, and the necessity of using lengthy periphrases. Such expressions as *put a question* and *have a desire for* make the language “inevitably rather neutral or slightly formal” (Carter, 41).

According to Nyberg et al. (2003, 250), Basic “was seen as a mere curiosity, unsuitable for any practical purpose”. However, the system gained Winston Churchill’s acceptance, who regarded it to be “worth serious consideration” (McArthur 1998, 115). Even in spite of Palmer and West’s heavy criticism, who considered Basic as “a kind of second-class artificial pidgin English”, Churchill was not discouraged from believing in its practical usefulness (McArthur 1998, 115). As a result, he asked Palmer to develop Basic further in order to “make it more natural” (115). The idea that Basic would become an international medium of communication was especially appealing to Churchill. In a speech delivered at Harvard in 1943, he praised the idea of an international language:

Here you have, a deftly wrought plan for an international language capable of very wide transactions of practical business and of interchange of ideas ... It would certainly be a grand convenience for us all to be able to move freely about the world – and to find everywhere a medium, albeit primitive, of intercourse and understanding. (cited in Flesch 1944, 339)

In spite of the heavy criticism, the influence of Basic English on teaching English continues to be felt in modern times. Today the vocabulary makes part of Wikipedia’s Simple English and is used by language education companies, such as the Berlitz English Language School (Goddard and Wierzbicka 2018, 19). A number of literary, business, and scientific readers written in Basic English have been prepared as teaching materials, for example, *Robinson Crusoe*, *Julius Caesar*, *Stories from the Bible*, *Basic for Business*, and *Basic English Applied: Science*, to mention but a few (Battenburg 1994, 137). Simplified readers are now advocated as an effective form of extensive reading (Krashen 2011, 27). Because they are easy to read, they are more encouraging than authentic texts (Martinez 2015). Readers serve as a bridge to reading texts in original English. Basic English found its way into lexicography; for example, *The Basic Dictionary* (1932) renders the most common 7,500 words of “normal English” into Basic English, and *The General Basic English Dictionary* (1940) offers around 20,000 words and their meanings defined in Basic.

Basic English was intended to satisfy the fundamental needs of international communication, especially in commerce, radio, and science. Though the idea of simplified English was not accepted immediately, it provided inspiration for

broadcasting services. Voice of America (VOA) has been broadcasting radio programs in Special English since 1959 (VOA *Special English Word Book A–Z* 2009). Special English is a simplified language for communicating with learners of English around the world. It uses a limited vocabulary of around 1,500 words, which is largely the same as that of Basic. The service broadcasts programs on radio stations, television, and, more recently, the Internet. It presents news, stories, and articles on a variety of topics. The transcripts and audio files can be downloaded and used for teaching listening comprehension at the intermediate level. The restricted vocabulary and the slow pace of speech make the texts accessible for less proficient learners.

As a consequence of the severe criticism of Basic, Ogden declined the invitation to the Carnegie Conference in 1934, the first conference devoted to the development of a general-purpose vocabulary (see 1.5.1). Ogden's refusal to participate in the conference prevented him from contributing to the shape of the vocabulary. Ogden had emerged as a leader in the field of vocabulary control. However, for West and Palmer, he seems to have been a rival, who in spite of his little experience as a teacher of English as a foreign language, was capable of exploiting his research in the production of teaching materials on a grand scale (Howatt 1984). Following his refusal, the decisions regarding vocabulary selection were left in the hands of the other researchers.

1.5 The General Service List: Confrontation of the competing approaches

The intense rivalry between objectivists and subjectivists was not resolved immediately, but the scholars took opportunities to confront their competing perspectives to create a common list. This section provides an account of how the approaches were challenged at the Carnegie Conference, and how the final vocabulary list took shape. The section also presents critical voices raised against the list.

1.5.1 The Carnegie Conferences

In 1934, a year before the publication of West's *NMED*, leading researchers in the field of vocabulary selection were brought together at the first conference devoted to this topic. The conference was held in New York on the initiative of West and under the sponsorship of the Carnegie Corporation. The aim was to investigate the role of word-lists in teaching English as a foreign language, and to develop a vocabulary that would simplify teaching, and serve as a basis for future work (Faucett et al. 1936, 1). In 1935 the Carnegie Corporation convened another conference in London to encourage further collaboration between the researchers.

A team of experts in the field undertook the task of developing the list. The committee in charge of the task included key figures of the vocabulary control movement: Lawrence Faucett, Harold Palmer, Edward Thorndike, and Michael West (Faucett et al. 1936, v). Among contributors was Hornby, who was responsible for preparing collocations. As mentioned earlier, Ogden did not attend the conference. The committee worked collaboratively and made decisions on the basis of the analysis of existing lists, namely the Thorndike-Horn list as compiled by Faucett and Maki in 1932, subjective lists prepared by Palmer and West, and several other “unidentified” lists (Faucett et al. 1936, 12; cited in Gilner 2011, 68). The researchers began with the examination of the 1,500 top-ranked words in the Faucett-Maki-Thorndike-Horn list. They marked doubtful words and discussed them individually, while keeping separate lists of words “desirable but open to reconsideration” and “words of possible utility to be reconsidered later” (Faucett et al. 1936, 11). However, no word was included in the final list “save on a unanimous vote” (11). Then, the vocabulary was revised by consulting subjective lists, among which was West’s *Defining Vocabulary* (1935). Next, the compilers examined words ranked up to 5,000 in the Faucett-Maki list, and discussed difficult points until reaching an agreement upon the final content. Although subjectivists and objectivists did not strive for reconciliation, there was a great overlap in the results of their studies as far as the first 1,500 words were concerned (Faucett et al. 1936, 13). However, beyond that boundary they found it difficult to select words that would satisfy both groups of scholars. The task of finding key words was even more difficult above the level of 2,500 because of numerous stylistic variants (13).

The Carnegie list was subjected to further revision at the second Carnegie Conference in 1935. The final vocabulary, which was comprised of approximately 2,000 items, was published as the fifth part of the *Interim Report* (1936). Although the list appeared under the heading the General Service List (GSL), for many years it was known to English teachers as the *Interim Report* (Jeffery 1953).

According to West (1953, vi), the publication of the *Interim Report* (1936) was the culmination of research on vocabulary control. In the post-war years, the interest in vocabulary selection became somewhat diminished, as other areas of linguistics, notably structuralism, attracted a great deal of scholarly attention (West 1953, vi). However, as we shall see, language scholars would persist in developing controlled vocabularies, and the GSL would have far-reaching consequences for language study and teaching.

1.5.2 The General Service List

The General Service List (GSL 1936) was compiled to satisfy the general needs of foreign learners of English and their teachers. In the selection of words, the following criteria were used (Faucett et al. 1936, 13):

- a. Word frequency. This criterion allowed for the inclusion of frequent words to the extent that they satisfied the subjectivists' decision. The compilers observed, however, that the results of the objective approach as far as the first 1,500 words were concerned corresponded closely to the subjectivists' judgements. Beyond that level the compromise was more difficult to achieve, but the criterion of frequency was still applicable. This was the case of words of "special stylistic value", which were more precise than semantically general alternatives; they were acceptable as long as they were of reasonable frequency (13).
- b. Structural value. Words of structural value comprised all function words, such as *at*, *for*, *to*, etc.
- c. Universality with regard to geographical area. This meant the inclusion of words unrestricted to any place or country, with the exception of words which, though not being widely current, were increasingly used in English, for example, *telephone*, *belt*, *scout*, *park*, and sports terms. Also it was acceptable to use those American variants which "appeared desirable as simplifications" and which had some sanction in England, for example, *airplane* instead of the British *aeroplane*, and *bucket* and *vine*.
- d. Subject range, or the applicability of words to a range of subjects. For example, the compilers included *bill* because it applied to "the general purchasing public", while excluding *invoice* because it was used only by "the seller or business man" (15). Examples of words which failed to satisfy this criterion were *essay* and *blackboard*, because they were of no use for people who have left school. As for words for professions, *lawyer* was acceptable, but *solicitor* was not.
- e. Value for defining purposes. The compilers selected three types of words needed for a learner to provide and understand a definition. The first type included "words necessary to enable a pupil to satisfy a need which he is unable to express directly"; the second, "words necessary to enable a teacher to define an unknown word without use of the vernacular"; and the third, "words necessary for understanding a dictionary" (17). The compilers drew a line between general-purpose words used in definitions and more precise but technical words (such as *fungus*) which are necessary for defining other words. They suggested that the latter be relegated to a supplementary list, rather than be part of the main body of the GSL.
- f. Value for word-building. The fact that both *in* and *ject* are meaningful justifies the inclusion of *inject*. By the same token, *reject*, *import*, and *export* are acceptable.
- g. Stylistic function, or the suitability of a word for different discourse types. The compilers excluded colloquial words, such as *awfully* in the meaning of "very", because they were of no use in written discourse. The compilers

generally avoided synonyms for general words, and derivatives which could easily be confused with other similar words, for example, *eventual*, for it could be interpreted as “as it may turn out”, and *fatal*, which could be confused with *fated* (18). The researchers admitted that any restricted vocabulary may result in vagueness and the loss of accuracy. In order to avoid this risk, they even attempted to make “an adequate provision of Words of Precision”, such as *thick* for width and *dense* for other meanings, but the plan was thwarted by “the conventions of the language” and impracticality of this strategy (18).

The above criteria juxtapose the objective measures of word frequency and range with the subjective criteria of utility for learners and definers.

Like *Thousand Word English*, the GSL resembles a dictionary more than a simple word-list. It has approximately 2,000 entries arranged alphabetically, with headwords, brief definitions, and extensive use of illustrative examples. Each capitalized headword provides access to the word family, which offers – in the editors’ words – “connected changes of form and meaning” (87). A word family includes generally transparent derivatives, inflected and alternative spelling forms, compounds, compound elements, and occasionally word combinations (e.g., *for fear that*, *in fear of*), all being gathered together irrespective of the part of speech. Senses⁵ are numbered. Below are two examples of the GSL entries.

BACK

- back**, n. (1) (*part of body*)
 backbone, n. I have hurt my back
 Turn one’s back to the audience
 Behind one’s back
 (2) (*opposite of front*)
 The back of the house
 The back of a knife

5. In metalexigraphy, the concept of *sense* refers to one of the basic divisions of a dictionary entry, which covers a definition and is distinguished by a number or letter (Piotrowski 1994, 21; Adamska-Sałaciak 2007). This meaning of *sense* is distinct from that which is used in structural semantics to talk of the set of semantic relations (in particular paradigmatic ones) that hold between a lexical unit and other lexical units in the language (Burkhanov 1998; Lew 2013, 284). For the purpose of dictionary analysis, it is more reasonable to understand *sense* in the way the term is used in linguistic semantics and lexicology, and treat it as a meaning of a polysemous lexical unit. However, caution needs to be exercised against the indiscriminate equating of lexicographical senses with lexical ones. The difference between the two tends to be ignored by dictionary users, who believe that dictionary senses “mirror what goes on in the language” (Lew 2013, 285). Suffice it to say that lexicographers do not agree on such matters as how many senses a word has and how the sense boundaries should be determined.

- The back garden
 The back door
back, adv. prt. (1) (*to the rear*)
 Go back
Phrases:
At the back of
Keep back
 Keep back a fact (*conceal*)
 Keep back money (*not pay*)
 (2) (*returned to first position*)
 Is he back yet?
 Go there and back
 Go, come, run, back
 Put, pull, push, throw, something back
 Pay back
background, n. (*of picture, mind or life*)
 Keep in the background
backward/s, adv. (1) (*away from front*)
 Go backwards
 Backwards and forwards
 (2) (*reverse order*)
 Say the letters backwards, Z Y X ...
- SHIP**
ship, n. A good ship
shipment, n.
shipwreck
 (?) [*shipping* = *ships*]

As a general rule, when selecting derivatives for a word family, the researchers took into consideration their close relationships with the base form and the amount of effort required for the recognition of these relationships. They highlighted semantically transparent derivatives by italic type in order “to give some idea of varying degrees of learning effort” (87). On the other hand, *hard* and *hardly* figure as separate word families, and so do *family* and *familiar*, because of weak semantic link between them (see also Bauer and Nation 1993).

Occasionally, the editors admitted lexical items which fell outside the boundaries of the general vocabulary, but indicated their special status by inserting them in square brackets, as in *shipwreck* above. The question mark that precedes the brackets indicates that, according to the editors, “it is doubtful whether this item is worth teaching” (West 1953, ix). This shows that the compilers were sometimes liberal in their selection, the point for which they were criticized (see 1.5.3).

After the *Interim Report* was published in 1936, arrangements were made to revise the GSL, and it was West who was entrusted with this task (Jeffery 1953, vi). However, the progress of the work was delayed because of the outbreak of the Second World War, and it was not until 1953 that the revised word-list was published under a similar title: *A General Service List of English Words* (GSL).

A great improvement on the *Interim Report* was the introduction of the information on the frequency of words and meanings. This information was prepared by two American researchers: Irving Lorge and Edward Thorndike. Their contribution to the project was immense, given the fact that their counts were the product of a manual exploration of a large corpus, which originally contained 2.5 million words, and later doubled in size (Lorge 1953, xi). Because the corpus was being expanded in the course of the project, some counts provided in the GSL referred to the earlier version of the corpus.

As shown below, each entry provided the frequency of occurrence of the headword (or sub-head) in the corpus and the percentage of occurrence of each sense of the headword. The entry for **game** shows that the lexeme occurred 638 times, with the sense of “amusement” constituting 9% of the total frequency, the sense of “indoor and outdoor games”, 38%, and so on.

GAME	638	(1) (<i>amusement, children's play</i>)
Fun and games		
It's not serious; it's just a game	9%	
(2) (<i>with the idea of competition, e.g. cards, football, etc.</i>)		
A game of football		
Indoor games; out-door games	38%	
(3) (<i>a particular contest</i>)		
We won, six games to three		
I played a poor game		
Play a losing game (10.5%)	23%	
(4) (<i>Games = athletic contest</i>)		
Olympic Games	8%	
? [= <i>animals</i> , 11%; <i>game-/</i> , <i>game-birds, etc.</i> , 5%]		
[= <i>fun</i> , <i>Make game of</i> , 0.5%]		
...		(West 1953, vii)

Lorge and Thorndike had reported semantic counts in their earlier work, *A Semantic Count of English Words* (1938). In 1949 Lorge prepared a supplement to this report, under the title *The Semantic Count of the 570 Commonest Words*. These resources, after rearrangement, served as ready-made data for the use in the revised version of the GSL. In Lorge (1953, xi), we find a brief description of the procedure of semantic count. As the compiler explains, the sample was drawn from various sources, such as encyclopaedias, magazines, novels, coursebooks, essays, biographies, scientific and literary texts. The researchers engaged specially

trained assistants to read through the corpus. They were responsible for the identification of meanings of all words in the corpus and of making notes on their location by page number, unit, and column. The assistants identified the meanings by reference to the *Oxford English Dictionary* (the 1933 edition). As soon as they found a word in the corpus, they matched its meaning to the corresponding sense in the *OED*. The word was then marked with its meaning and sense symbol as given in the *OED* (xii). By providing frequency information on form and meaning, the GSL became a highly informative resource for course designers.

Although the GSL (1953) is generally considered a reissue of the Interim Report, West rearranged a number of meanings according to Lorge's semantic data. While comparing word meanings recorded in the *Interim Report* with those in the *OED*, it was found that the former sometimes included the meanings which were either rare or unattested in the *OED*, and it even failed to record some fairly frequent meanings (Lorge 1953, xiii). For example, the meaning of *game* "have the game in one's hands", which was recorded in the Interim (see the entry **game** below), occurred with a very low frequency (twice in a thousand), whereas the meaning "animals used for sport" occurred far more frequently (xiii). These facts led the researchers to modify the content of the entry accordingly, by removing the former rare meaning, and retaining the latter (albeit with a question mark indicating that this meaning is still undesirable for teaching purposes). Comparing the 1953 entry for **game** (cited earlier) with the corresponding one from the Interim Report (1936), one finds more differences. Firstly, the new entry distinguishes between several meanings of the word and does so by means of definitions, rather than just illustrative examples; secondly, by providing frequency information, it highlights those meanings which are worth teaching.

GAME

game, n. A game of football

Play a losing game

Have the game in one's hands

(x) [*game* = animals used for sports; *game* = tricks;

have a game with = make fun of]

(Faucett et al., 1936)

Worthy of mention is the fact that West marked some changes explicitly as additions. For example, in the entry for **line**, he inserted a comment on the verbal use of the word (Gilner 2011):

This word is not in the Report ... but it covers an idea which can hardly be expressed otherwise:

A coat lined with silk, with a silk lining 2%

(West 1953, 280)

In the revised GSL (1953, ix), West enumerates, besides word frequency, several factors that are worth considering in vocabulary selection for teaching. They are a reflection of the criteria used in the preparation of the 1936 version of the list:

- a. Frequency. The criterion of frequency is useful in the selection of both words and their meanings.
- b. “Ease or difficulty of learning”. This factor reflects the amount of effort needed by a learner to learn a new word or meaning. According to West, some meanings are so closely related to each other that learners who know one of the meanings can acquire another effortlessly. This factor, however, should be counterbalanced by other factors; for example, the relation between two meanings of *claim*: “mining” and “claim the title” is close, though the former one is rare (ix).
- c. “Necessity”. A word that expresses a “certain range of necessary ideas” is worth including even though it occurs with a relatively low frequency. For example, the verb *preserve* in the sense of “food” is such a word, because it is “the only satisfactory cover for canning, bottling, salting, freezing, jam making” (West 1953, ix).
- d. “Cover”. The notion of *cover* should be understood as the reverse situation to that explained in point (c) above. A lexical item may be frequent but unnecessary, because there is another item which is already in use, and whose meaning covers the former. For example, it is not desirable to include *for the time being* because it can be replaced with the synonymous expression *for the present* (ix).
- e. “Stylistic level”. A learner does not need literary words (*personage*), nor stylistically low words (*fellow, chap*), but “should keep a middle path” (*person*) (x).
- f. “Intensive and emotional words”. These items are “of secondary importance to the foreign learner”, because the primary aim of learning English is to “express ideas rather than emotion” (x).

Using Cruse’s terminology (2004), we would now say that West’s criteria (e) and (f) gave priority to descriptive rather than non-descriptive dimensions of meaning, and to words representing basic level categories (see also Ungerer and Schmid 2006, 329).

The GSL (1953) came out with an appendix titled “Supplementary Scientific and Technical Vocabulary” and compiled by W.E. Flood and M. West. The supplement was intended for readers of specialist texts, who had not received training in scientific or technical fields (Flood and West 1953, 583). The vocabulary was supposed to be comprehensible for non-specialist readers. It is comprised of 425 words arranged in alphabetical order with no word class labels. Sometimes

the form of a headword reveals that two word classes are involved, for example, GENERATE/OR. As the compilers explain, “the adequacy of the vocabulary” has been tried out by paraphrasing extracts from popular scientific books (584). The compilers used most of the supplementary words, except for so-called “wanted” words (see below), in the defining vocabulary in *An Explaining and Pronouncing Dictionary of Scientific and Technical Words* (1952). The supplement consists of the following categories of words (583):

- a. “Essential scientific words often needed”, such as *atom, electrolysis, protein, and valve*;
- b. “Scientific words of moderate necessity but unavoidable when required”, for example, *condenser, neutron, and violet*;
- c. “Key words which facilitate the explanations of groups of ideas”, such as *crank, lever, piston, and yeast*;
- d. Words excluded from the GSL, for example, *ape, bee, jute, and coconut*;
- e. “Dignity” words, that is, the ones which “could have been avoided” but were “retained to avoid childishness and over-simplification”. Examples are: *consists of, method, similar*, which were preferred to their simpler alternatives: *is made of, way, and rather like*, respectively;
- f. “Wanted” words, that is, the ones which the compilers have not used in the dictionary, but they recommend their inclusion in scientific word-lists because they are difficult to avoid. They include nouns, such as *abdomen, carrot, filter, fuel, owl, pear, pedal, truck, tunnel*; a few verbs, including *to correspond to, to dilute, to spray*; and some adjectives, for example, *concentrated, equivalent, and temporary*;
- g. Words considered international, for example, *beer and cigarette*.

It is interesting to see that a large number of words included in Flood and West’s supplement but excluded from the GSL represent the natural kind category (Stein 2002b). It may be that such words are particularly difficult to paraphrase. Stein distinguishes three subcategories of these words (Stein, 184):

- a. Names of body parts: *ankle, bowel, breast, calf, cheek, chin, kidney, liver, muscle, nerve, nostril, pupil, rib, scale, skeleton, skull, spine, spinal cord*
- b. Names of animals: *ant, ape, bat, bedbug, bee, beetle, butterfly, crab, flea, frog, hen, lamb, lizard, mosquito, moth, owl, poultry, snail, spider, toad, whale*
- c. Names of plants and their parts: *bud, bulb, cabbage, carrot, coconut, daisy, fern, fibre, fir, flax, gourd, grape, jute, lemon, lily, maize, moss, onion, palm tree, pear, petal, plum, pod, pollen, potato, rose, sponge, vegetable*

Flood and West arguably believed in the explanatory power of their vocabulary, and considered it indispensable in defining scientific and technical terms (Stein 2002b). As they claimed, the vocabulary combined with the GSL was “capable of expressing all ordinary scientific and technical subjects within the limits of non-specialised study” (Flood and West 1953, 583).

The compilation of the GSL had far-reaching consequences for language teaching. Until the 1970s, the GSL was considered “the accepted and indeed the only objective authority on what to include in syllabuses for the teaching of L2 English” (Widdowson 1968, 123; cited after McArthur 1998, 67). In Section 2.2, we will see that the influence of the GSL in EFL teaching would extend beyond that period, even in spite of the criticism of the list.

1.5.3 Criticism of the GSL

The GSL has been criticized, among others, by Engels (1968) and Richards (1974) for insufficient range and datedness. However, according to Gilner (2011), this criticism is not entirely justified. Engels (1968, 226) reports a study examining how the GSL vocabulary (3,372 items⁶) was distributed in 10 randomly selected texts of equal length (1,000 words each). The researcher set out to check whether the GSL words qualified for general vocabulary by examining whether the words were distributed widely enough in these texts. Engels made an alphabetical list of all the running words that appeared in the texts and matched them with the corresponding words in the GSL. The GSL words were ranked according to the frequency with which they occurred in the texts. He found that while the first 1,000 top-frequency words from the GSL occur in up to 10 texts, the highest range reported for the words ranked 1,000–2,000 is 5 texts, and for the words ranked 2,000 – 3,000, only 2 texts. Engels argues that the words ranked below 1,000 do not have a sufficiently wide distribution across the texts, and thus cannot be called “general service-words” (Engels, 226). However, as Gilner (2011) observes, this study is flawed by the small size of the text sample (10,000 words in total) and the unrealistic expectation of finding a GSL word in each section of the corpus. As he remarks,

[e]ven if one were to think up ten 1,000-word texts made exclusively from the 3,372 GSL words and having these words uniformly distributed across the texts, each word could not appear in more than three texts. [...] it is numerically impossible to find 3,372 words in a 1,000 word text. (Gilner 2011, 71)

6. This number apparently indicates the total of headwords and members of the word families, including part-of-speech homographs.

In turn, Richards (1974, 71) claims that the GSL “contains a great number of words of limited utility” because they are used to talk only about a limited range of topics. Examples of such words are *apologize*, *curse*, *crown*, *express*, *fear*, *fade*, *ornament*, *scent*, *mild*, *motion*, and *nest*. According to the researcher, one of the reasons for their limited use is that they have become outdated, as the GSL was based on the word counts from the early decades of the 20th century. Therefore, the scholar suggests updating the list by adding the words in common use in the 1970s, such as *astronaut*, *helicopter*, *pilot*, *rocket*, and *television* (Richards, 71). It is worth adding that Richards’ reservations about the outdatedness of the list are shared by Carter (2012, 198), who cites the same words, but *rocket*, as being “common” in the 1980s.

However, as Gilner argues (2011, 71), the above claims are not supported entirely by the evidence from the BNC. The argument is that some of the words of allegedly limited utility are roughly similar in terms of three parameters: frequency, range, and dispersion,⁷ to those which Richards suggests for inclusion. Therefore, according to the BNC frequency list (Leech et al. 2001), the values of the parameters for the two sets of words are as follows:

apologise 11, 96, 89

mild 18, 100, 93

fear 55, 100, 94

scent 11, 88, 88

astronaut 2, 50, 81

helicopter 16, 96, 88

pilot 43, 100, 88

rocket 9, 92, 88

Thus, the words in the upper set of words, which Richards suggests for inclusion, occur in the BNC with low frequencies (11, 18, 55 and 11 respectively per million word tokens), but their values of range and dispersion are high, reaching the maximum level. These results are roughly comparable to those in the lower set (Gilner 2011, 71).

7. Leech et al. (2001) used these criteria to describe the lexical content of the BNC. Frequency value means the frequency per million word tokens; range is the number of sectors of the corpus in which a word occurs (max. 100); and dispersion (max. 100) is a statistical measure of how evenly a word is distributed across successive million-word sectors of the BNC (Leech et al. 2001, 18). Words may be frequent either because they occur in a number of samples of the corpus or because they are often used in only a few samples. Dispersion helps us distinguish between the two situations. A frequent word with a high dispersion value has high usage in the corpus as a whole (Leech et al., 18).

Some recent studies call for revision of the GSL. They demonstrate that the GSL shows vocabulary gaps and gives undue emphasis to the areas of lexis unfamiliar to the average learner. Browne (2014, 37) remarks that the GSL is rich in the areas of vocabulary which no longer reflect the needs of the modern EFL learner; for example, “nautical terms (*oar, vessel, merchant, sailor*, etc.), agricultural terms (*plow, mill, spade, cultivator*, etc.), religious terms (*devil, mercy, bless, preach, grace*, etc.)”. Other questionable terms, which have long been out of date, are *telegraph, coal, gaiety* (in the meaning of “happiness”), *shilling, servant, footman, milkmaid* (Browne 2014; Brezina and Gablasova 2015). Another such word is *porter*, which by the 1980s was no longer in common use in the UK (Michael Rundell: personal communication). On the other hand, the GSL omits terms for technological innovations which have only recently become widespread, for example, *computer* (Nation and Hwang 1995). Note that the values of frequency, range, and dispersion for this word are high: 174, 100, 85, respectively (Leech et al. 2001).

Claims in connection with the selection of words within lexical fields were already raised by Richards (1974, 71). He questions the inclusion of *trader, merchant*, and *dealer*, arguing that only one of them would suffice as a substitute for the others. In turn, Carter (2012, 198) points to the fact that the GSL, being based on a written corpus, has a number of words which sound “distinctly literary”. Nevertheless, he admits, their inclusion is justified, as the list was intended for the preparation of reading materials. Moreover, some critics have pointed to the inconsistent selection of derivatives; for example, under *effect* one finds an unrelated word *affect* (Bauman 2018a). The inclusion of inflected forms together with derived ones under a headword prioritizes form not meaning. Finally, as Gilner and Morales (2008, 1) claim, the GSL has limited expandability, which means that it is virtually impossible for a researcher to expand the vocabulary according to purely “the original directives and criteria” (Gilner and Morales 2008).

All in all, it is hard to disagree with the fact that the GSL calls for amendments and corrections. Taking into consideration its age, it may come as no surprise that researchers express doubts about the relevance of the list, though they do not always provide sufficient evidence to support their claims. Yet in spite of the above shortcomings, some scholars not so long ago considered the GSL as “the best available” in language pedagogy (Nation and Waring 1997, 15) and “essential for all learners no matter whether they are using English as a foreign or second language, for spoken or written use, or for general or special purposes” (Nation and Hwang 1995, 36). The reasons for such opinions will be presented in Section 2.2.

1.6 Summary

In this chapter, we have seen that although the use of a restricted defining vocabulary in *NMED* (1935) was completely novel to English lexicography, this innovation owed a great deal to the vocabulary control movement of the 1920s and 1930s. The movement manifested itself in the practical and scholarly interest in vocabulary selection pursued by language researchers and educators. Their common aim was to lighten the burden of learning a foreign language by restricting the vocabulary to a necessary minimum. The GSL was the culmination of the collaborative efforts of researchers. The list served course designers and teachers as lexical bases for coursebooks, simplified readers, and other teaching material. Although the interest in vocabulary control somewhat waned in the years following immediately the publication of the GSL, a number of other vocabulary lists were developed towards the turn of the 21st century. They will be the subjects of the following chapter.

Other major projects on vocabulary control

The pioneers of the vocabulary control movement directed their efforts towards developing vocabularies for general pedagogical and communication purposes. After the publication of the GSL in 1953, this line of research went into eclipse for around 20 years until the computer revolution, which opened new opportunities in vocabulary control. This chapter focuses on major projects undertaken since the 1970s which led to the creation of other restricted vocabularies. Some of the projects had clear pedagogical aims (see 2.1), others were started to enhance professional communication (see 2.3). In Section 2.2, we will compare the pedagogical lists with the GSL, and discuss some reasons for the long-lasting popularity and usefulness of West's list. In Sections 2.4 and 2.5, we will take a closer look at the concept of core vocabulary by considering theoretical, lexicographical, and pedagogical issues relevant to this concept. Section 2.5 discusses the implications of Wierzbicka's Natural Semantic Metalanguage (NSM) for vocabulary control in lexicography.

2.1 Word-lists for pedagogical purposes

It is impossible to describe all the word-lists which have been created and used in language education since 1953, and the interested reader may find useful reviews in Bongers (1947), Stein (2002b), Tickoo (2008), and Youngblood and Folse (2017). The survey below is by necessity limited to those lists which are the most influential and noticeable in the pedagogical literature. This selection includes word-lists intended for learners of English as a foreign and second language (EFL and ESL, respectively)¹ as well as for native speakers of this language. Some of them have

1. The terms *foreign language* and *second language* designate distinct concepts. A second language plays an important role in the learner's new country of residence where it is widely spoken. For example, immigrants living in the US learn English as a second language. English has also a privileged status in India, Nigeria, and Singapore, where natives of these countries learn English as a second language because it is the main language of business, education,

been designed as general-purpose lists, and others as academic (special-purposes) vocabularies.

2.1.1 The American Heritage Word Frequency Book (1971)

The American tradition of compiling word-counts for childhood education initiated by Thorndike and Horn continued into the second half of the 20th century (Landau 2001). In 1971, John Carrol, Peter Davies, and Barry Richman created *The American Heritage Word Frequency Book*. Like Thorndike's vocabulary, the list was based on frequency and range, but unlike its predecessor, it was compiled with the aid of a computer. The corpus was composed of 5 million words culled from over 1,000 various written sources, mainly textbooks used in elementary and junior high schools in the United States (Landau 2001). The list was comprised of high frequency words occurring across a range of grades (between 3 and 9) and subject areas (Nation and Waring 1997). It is the distribution of words across various sections of the corpus that is praised as a valuable feature of this list (Lovelace 1988). The *Word Frequency Book* was used in the creation of the *American Heritage School Dictionary* (Landau 2001).

However, critics point out that the *Word Frequency Book* was intended to represent "the printed language of the American elementary educational system" rather than the language of the actual books and other reading materials to which children were normally exposed (Carrol et al. 1971, as cited in Landau 2001). Therefore, as Landau (2001, 284) remarks, the book is a reflection of the "past assumptions of previous textbook writers about the vocabulary suitable for children of various ages". This fact makes the list more useful for textbook historians and researchers than for course designers. Nevertheless, the *Word Frequency Book* is one of the early advances in computerized analyses of texts with the use of statistical techniques (DeRocher 1973).

2.1.2 The Threshold Level English (1975)

In 1975, J. van Ek designed the Threshold Level English for the Council of Europe. The driving force behind the project was the need to foster closer relations among European countries regardless of differences in culture and heritage, and to improve learners' command of English as a foreign language (Ek 1975; see also

government, etc. (Richards and Schmidt 2010). By contrast, a foreign language does not have that status in the learner's native country, and is not used commonly there. It is normally taught as a school subject in order to enable learners to communicate with foreigners. The two terms are sometimes used interchangeably, especially in the US (Crystal 2003), but they are intuitively distinguishable from *native language*, or *mother tongue*.

Stein 2002b). The Threshold Level was developed as a system of credits in order to unify the requirements for adult language teaching in various educational institutions across the European countries. The designer adopted a definition of *threshold level* as “the lowest level of general foreign-language ability to be recognized in the unit/credit system” (Ek 1975, 8). The general ability allows learners to “maintain themselves in most everyday situations” (Ek, 8), especially when they visit foreign countries or come into “temporary contacts with foreigners in their own country” (Ek, 9). The threshold level specifies settings, situations, topics, and concepts in which a learner is likely to be found.

The vocabulary was developed from scratch. The first step was to list relatively universal concepts that express properties, qualities, location, motion, time, events, etc. Then the compilers determined “the absolute economical minimum of lexis” which the learner should be familiar with in order to talk about these concepts in a given language (Whitcut 1988, 49–50). This methodology is reminiscent of Ogden’s approach to the compilation of Basic English over 40 years earlier. Having identified the vocabulary used in real-life situations, Ek ended up with a list of 1,500 lexical items. About two thirds of the list were selected for both receptive and productive purposes, while the rest was meant for receptive use only. These distinct uses were indicated by labels P and R, respectively. The vocabulary is arranged in alphabetical order, and the majority of the items are flanked by sentences illustrating their typical usage.

The need to establish further levels of language proficiency led to other projects. In 1977, J.A. van Ek in collaboration with L.G. Alexander and M.A. Fitzpatrick published *Waystage English*, an elementary course for adults (Ek et al. 1980; see also Stein 2002b). The authors defined *Waystage* as a level of communicative competence “halfway between zero and the threshold level” (Ek et al. 1980, 2). Towards the turn of the century, both *Threshold* and *Waystage* were being revised and further extended (Ek and Trim 1990a/1998a, 1990b/1998b, 2001). In 2001, they became part of a broader system, called the *Common European Framework of Reference for Languages* (CEFR). CEFR specifies six degrees of competence, from beginner to proficiency: A1, A2, B1, B2, C1, and C2. These labels are used by some dictionaries (e.g., *CALD4*) to indicate which words should be learnt at the designated levels of proficiency. The system is now widely used by educational institutions across Europe as a basis for foreign language course planning, assessment of learners’ achievements, and recognition of language qualifications.

2.1.3 The Cambridge English Lexicon (1980)

The Cambridge English Lexicon was designed as a receptive vocabulary that would be sufficient for EFL learners to pass the Cambridge First Certificate in

English (FCE). It was compiled by Roland Hindmarsh in 1980, who had been commissioned to undertake this task by two examination administrators: The British Council and the Local Examinations Syndicate of the University of Cambridge (Stein 2002b). Hindmarsh drew on a number of earlier word-lists, including Ogden's Basic English (1930), Thorndike and Lorge's *The Teacher's Wordbook of 30,000 Words* (1944), West's GSL (1953), and Kučera and Francis' list (1967) (Stein 2002b). The last list mentioned will be discussed in Section 2.1.4.

The *Lexicon* consists of 4,500 words with more than 8,000 senses specified. As the compiler claims, it is particularly suited to "text editors, curriculum planners, examination setters and course writers at or below the level of reading comprehension required for the First Certificate in English" (1980, xii). A feature that makes the list particularly useful for course designers is that the words are graded according to frequency, from 1 to 5 (Gairns and Redman 1986). The grades provide approximate guidance on the level at which a word should be introduced on a course leading to the FCE. Thus, words marked by grade 1 are frequent in English and are suitable for beginners, while those graded 5 are for students preparing for the FCE exam (Gairns and Redman 1986, 58). The extra grades 6–7 indicate the vocabulary appropriate for higher levels of English proficiency. In addition, the *Lexicon* provides senses, which are also graded. Hindmarsh (1980) does not specify how the senses have been selected or graded, but he mentions two dictionaries as the main source of that information: West's *An International Reader's Dictionary* (1965) and *OALD3* (1974).

The word-list is supposed to be treated as an approximate guide, rather than a set of strict rules, on what to include in a course. Thus, according to the compiler, it is justified to use a low-frequency word that expresses a concept essential for a particular communicative situation, but that cannot be replaced easily with a more frequent word (Gairns and Redman 1986). Such less frequent words are unavoidable in specific situations. However, learners have different lexical needs, depending on the mode of communication (spoken vs. written). Words like *sweater*, *jumper*, and *pullover* may be useful for reading comprehension, but the knowledge of any one of them will suffice for productive purposes (Gairns and Redman 1986, 59).

2.1.4 Word-lists from popular corpora: the Brown (1967), the LOB (1978), the Bank of English (1991), and the BNC (1990s)

The American Brown Corpus (Brown) and its British counterpart, the Lancaster-Oslo-Bergen Corpus (LOB), are long-established resources for linguists interested in frequencies and distribution of words. The corpora owe their development to the advances in computer technology in the 1960s and 1970s, enabling researchers to store, organize and retrieve language data more efficiently than

ever before. The Brown and LOB corpora were compiled independently through a computational analysis of approximately 1 million running words. Developed in the 1960s by Henry Kučera and W. Nelson Francis from Brown University, the Brown corpus is a balanced collection of American written texts published for the first time in 1961 (Gove 1972), subdivided into 500 samples of equal length (approximately 2,000 words each). The samples represent 15 genres, including the language of the press, religion, belles lettres, and fiction (Kučera and Francis 1967). The LOB is similar in size and structure but differs from the former mainly in that it is comprised of British texts (Johansson 1978).

The above two corpora have proved invaluable for linguists in the studies of word frequency (e.g., Kučera and Francis 1967; Francis and Kučera 1982; Kjellmer 1986; Oakes 2009b). They were used by applied linguists, notably Nation and Hwang (1995), to identify a general service vocabulary which would be essential for foreign language learners in “most uses of the language” (Nation and Hwang 1995, 35). From each corpus, Nation and Hwang retrieved a list of the words which were highly frequent and widespread in the language. The words selected occurred in at least 10 out of 15 sections in each corpus. In this way, Nation and Hwang (1995) determined 1810 words in the LOB corpus and 2410 words in the Brown corpus. The word-lists thus obtained were further compared against the GSL, and the words overlapping the three lists qualified for inclusion in the general service vocabulary (for comparison of the lists see Section 2.2). The creation of large-scale corpora of English towards the end of the century paved the way to more elaborate word lists.

The 1980s and 1990s witnessed the compilation of large corpora designed specifically for learner’s lexicography. In 1980 the University of Birmingham and the publisher William Collins launched a research project named COBUILD. The aim was to compile a dictionary of current English on the basis of a large corpus of texts. A team of experts in lexicography and computing worked under the direction of John Sinclair. In the year of publication of the *COBUILD* dictionary (1987), the corpus amounted to 7.3 million words, which was more than 7 times larger than the LOB and Brown corpora (Renouf 1987). Because the corpus was compiled primarily for EFL lexicography, it aimed to represent “the type of language which many learners of English undergoing formal instruction traditionally encounter” (Renouf 1987, 16). It included EFL coursebooks, recent general texts from several major varieties of English (British, American, Indian, South African, Australian), and popular adult literature, while excluding poetry, drama, and historical as well as technical texts. Since 1987 the corpus has been constantly expanding, reaching 645 million words in 2006. In 1991, it was renamed *The Bank of English*.

From the 1980s onwards, the COBUILD corpus assisted lexicographers and researchers in collecting data relevant to dictionary creation. Most importantly, it

served as a database for generating concordances. The concordances were a genuine innovation in dictionary compilation (Atkins and Rundell 2008), as they provided the lexicographers with easy access to empirical evidence for word usage. The concordances were used as the source of illustrative examples for the dictionary. Researchers used the corpus to retrieve frequency data. For example, Fox (1989) drew a frequency word list to examine the defining vocabulary of the *COBUILD Essential Dictionary* (see 4.3). The corpus was used by the COBUILD lexicographers in dividing words according to their frequencies in the language. On the basis of frequency lists, the dictionary highlights common words with diamond symbols to draw the learner's attention. This information is intended to inform the learner as to which headwords are more important than others.

In the 1990s, other dictionary publishers were actively engaged in developing their own corpora. Cambridge University Press assembled the Cambridge International Corpus to help in the creation of, *inter alia*, the *Cambridge International Dictionary of English* (CALD 1995). On the basis of this corpus, the compilers of the dictionary extracted a list of high-frequency words and determined the frequencies of occurrences of their meanings. The list was used to create the CALD defining vocabulary (CALD2, vii). Likewise, Longman publishers compiled a corpus of American English to inform lexicographers in creating the second edition of *Longman Dictionary of American English* (1997). Its British counterpart, the *Longman Dictionary of Contemporary English* (LDOCE3/1995), used the British National Corpus (BNC), a product of collaborative work of several publishers and institutions (see below). These corpora served as sources for retrieving frequency data for Longman dictionaries. For example, the LDOCE3 lexicographers identified "the 3000 most frequent words in both spoken and written English" and highlighted these words in the dictionary with special symbols ("S" in the spoken and "W" in the written corpus) to draw the learner's attention (LDOCE3, vii–xiii).

Designed in the early 1990s, the BNC consists of over 100 million words. It is made up of 90% written texts and 10% spoken texts, both representing the British variety of English. Using the BNC data, Nation (2006) compiled fourteen word family lists arranged according to word derivation, range, and frequency in order to determine the size of the vocabulary that the foreign learner needs for comprehension of written or spoken English. In compiling the lists, priority was given to the words that are highly frequent and widely distributed in the language, on the assumption that such words are learnt before narrow-range and low-frequency words (Read 1988; Laufer et al. 2004). The words related by derivation in a transparent way found their way into the top lists on the grounds that they would not pose problems for the learner who knows word-formation rules. Nation's original 14 lists were subjected to continual revision, and the 2011 version consists of 25

lists (Nation 2012). In the current research, I will use these 25 lists for the evaluation of definition vocabularies (see 4.1.3).

The fundamental question for Nation was the amount of text coverage “needed for adequate comprehension to be likely to occur” (Nation 2006, 61). Text coverage is the percentage of word tokens in a text covered by (or present in) a particular word list (Nation and Hwang 1995; Nation 2004). Having found that with a text coverage of 98% most readers gain adequate comprehension of fiction (Hu and Nation 2000), Nation set out to determine how many word families of his lists covered the above percentage of the vocabulary in unsimplified and simplified novels, newspapers, graded readers, and children’s movies (Nation 2006). Among the texts analyzed were *Lord Jim* by Joseph Conrad, *Lady Chatterely’s Lover* by David H. Lawrence, *The Great Gatsby* by Francis S. Fitzgerald, newspapers from the LOB and Brown corpora, a simplified version of *The Picture of Dorian Gray* by Oscar Wilde, a popular movie for children *Shrek*, and spontaneous unscripted spoken English from the Wellington Corpus of Spoken English² (Nation 2006). Using the RANGE program (Nation and Heatley 2002), the researcher concluded that the learner must need the vocabulary of the first 8 to 9 word lists (8,000–9,000 word families) in order to understand what is being read.

2.1.5 The University Word List (1984)

The University Word List (UWL) was proposed by Xue and Nation in 1984 as a “vocabulary of university study”. It is a list of 836 word families which typically occur in academic texts, with text coverage estimated for around 8.5% in this type of discourse (Johansson 1978; Nation 1990; Bauman 2018b; Coxhead 2000). The UWL vocabulary is divided into 11 levels according to frequency and range. It is intended for students who have “studied English for several years at school”; therefore, it is assumed that they already know the core vocabulary (Xue and Nation 1984, 216). The UWL excludes those words that belong to the GSL, which makes the former list complementary to the latter.

The UWL was prepared by editing and amalgamating four earlier vocabularies prepared by Campion and Elley (1971), Praninskas (1972), Lynn (1973), and Ghadessy (1979). Campion and Elley’s list, which provided the major source of data, had been published by the New Zealand Council for Educational Research

2. The Wellington Corpus of Spoken English comprises one million words of New Zealand English compiled between 1988 and 1994. It consists of 551 2,000-word extracts of formal, semi-formal, and informal speech. The corpus includes all categories of monologue, parliamentary debates, interviews, and private conversations (Xiao 2008).

(Xue and Nation 1984). It served as part of a university entrance English language test for overseas candidates (the Language Achievement Test for Overseas Students) (Xue and Nation 1984). The aim of the list was to reflect the vocabulary to which students were exposed during university courses. In developing the list, the compilers examined 301,800 words from 23 textbooks, 19 lectures, and a selection of examination papers. The lectures were available in journals and dealt with 19 academic disciplines which enjoyed the highest enrolments in New Zealand universities (Xue and Nation 1984), including accountancy, anthropology, biology, chemistry, economics, and education. The compilers excluded all those general-purpose words that were found among the first five thousand words in Thorndike and Lorge's *Teacher's Word Book of 30,000 Words* (1944) (Xue and Nation 1984).

The second source for the UWL was designed by Praninskas and became known as the American University Word List (AUWL). It was intended for non-native students of English at the American university in Beirut. The aim was to help them follow university courses and understand what was being taught. The corpus for the AUWL consisted of samples of ten university textbooks on arts and sciences. From each book, Praninskas took all the words occurring on every tenth page and arranged them in alphabetical order. Because the focus was on academic vocabulary, the compiler eliminated general-purpose words which appeared on the GSL and other words which fell outside the scope of the project, that is proper names, abbreviations, and foreign words (Youngblood & Folse 2017). The compiler made a few adjustments to the list by conflating inflected forms under their base forms; removing hyphenated compounds and listing their components as separate elements; and changing negative forms (e.g., *illogical*, *motionless*) into the affirmatives (Stein 2002b). The final AUWL consisted of 507 base words and 840 forms derived from the base words.

The above two lists, compiled respectively by Campion and Elley, and Praninskas, were combined into a single list. Then the content of the list was further verified against Lynn's and Ghadessy's vocabularies. These vocabularies were made up of the words regarded as particularly difficult for foreign language students. In compiling them, students had been asked to write glosses above difficult words they found in their university textbooks. By locating the glosses, which were given in students' mother tongue, Lynn and Ghadessy identified difficult words (Xue and Nation 1984). Around 30% of the words from each of the Lynn and Ghadessy lists provided additional material for the combined list by Campion and Elley, and Praninskas. For more than 15 years (Gardner and Davies 2013), the list served teachers and course designers as a basis for the creation of university textbooks and other teaching materials, though it has been criticized for being compiled manually on the basis of small corpora covering a narrow and unbalanced range of topics (Coxhead 2000).

2.1.6 The Academic Word List (2000)

With the rapid development of computational tools and techniques towards the end of the 20th century, linguists called for a word-list that would be more representative of the academic discourse. In 2000, Coxhead proposed the Academic Word List (AWL), addressing the question of “which words are most worth studying” for university students (Coxhead 2000, 213). Compared to the UWL, the AWL was compiled under stricter principles, which explains the relatively small overlap between the lists (51%). The AWL designer made use of a much larger corpus of academic texts composed of around 3.5 million running words. The texts represent 28 subject areas divided into four sections: arts, commerce, science, and law. The AWL is made up of 570 word families selected according to the principle that a word family occurs at least 25 times in each section of the corpus (totaling 100 occurrences). The word families selected are those that are deliberately omitted from the GSL but are frequently used and widely distributed across academic texts. Among them are the families headed by *analyse*, *concept*, *research*, and *data* (Coxhead 2000).

The AWL provides 10% coverage of running words in a written academic text, which is slightly more than is covered by the UWL in the learned and scientific writings of the LOB corpus (8.5%) (Johansson 1978). Thus, although the AWL is significantly smaller in size than the UWL (570 vs. 836, respectively), it provides a similar coverage of academic texts as the latter list. In other words, it appears to be more useful for students than the UWL, as they need less effort to learn it with almost the same text covered (Coxhead 2000). It is worth noting that *OALD8* and *LDOCE5* draw learners’ attention to the AWL words by highlighting them with the label “AWL”.

According to Gardner and Davies (2013), the AWL superseded the UWL as a vocabulary standard in English language education until recent years. However, the researchers criticize the idea of building an academic list on top of the GSL. They find it questionable to assume that such a list should be comprised of words of lower frequency than the GSL. For one thing, the GSL is dated; for another, there is evidence that numerous AWL words belong to the most frequent words in the British National Corpus (Nation 2004; Hancioğlu et al. 2008; Neufeld and Billuroğlu 2005). In a similar vein, Gardner and Davies (2013, 5) report that 41% of the AWL word families are reflected by the first 2,000 lemmas³ of the Corpus of Contemporary American English (COCA). This

3. In corpus linguistics, lemma is used as a canonical form of a lexeme to represent all its inflected word forms in the same part of speech, for example, the lemma *write* stands for the verbs *write*, *writes*, *wrote*, *written*, and *writing*.

finding shows that a large proportion of the AWL vocabulary in fact constitutes a large subset of the most frequent words in (American) English. The frequent words included in the AWL are also the main reason for the high coverage of academic texts provided by the list (Gardner and Davies, 5). While the researchers emphasize that an academic word-list should contain high-frequency words, they doubt whether the AWL and the GSL reflect a clear division between the academic and the general vocabulary. For example, the GSL contains a number of high-frequency academic words, such as *capital*, *exchange*, *rate*, *account*, and *market* (Gardner and Davies, 5).

Gardner and Davies (2013, 3) also criticize the AWL, and other word family lists, for lack of a distinction between different word classes (e.g., *proceeds* as a verb meaning ‘continues’ vs. *proceeds* as a noun meaning ‘profits’), and for ignoring the fact that words belonging to a word family sometimes differ with respect to their core meaning (e.g., *react* ‘respond’ vs. *reactor* ‘device or apparatus’). All these problems make the word-family-based vocabularies questionable. What the researchers propose instead is the compilation of word-lists on the basis of lemmas, with grouping of inflected forms only.

2.1.7 The Common Core Vocabulary (2002)

Gabriele Stein in her book *Developing Your English Vocabulary* (2002b) emphasizes the role of paraphrasing in acquiring lexical knowledge by foreign language learners. She draws attention to the fact that vocabulary acquisition is one of the greatest problems for learners, especially at the intermediate level. She argues that the ability to paraphrase a lexical item for which a learner does not know the word can be learnt by mastering a defining vocabulary. Because paraphrasing is essential for the advanced command of a foreign language, the knowledge of the defining vocabulary can greatly help learners achieve that level of proficiency (Stein 2002b, 50). To put it another way, knowledge of a defining vocabulary will not only enable learners to understand words used in definitions but also provide them with “necessary tools to paraphrase” abstruse and rare words (Stein 2016, 760). Stein looks into the defining vocabularies of three EFL dictionaries and identifies the common vocabulary on the assumption that it “might be the most useful lexical minimum to allow foreign learners of English at the intermediate level to express themselves and to elicit from their interlocutors the names for whatever they have to paraphrase” (Stein 2002b, 52). The source dictionaries are *Longman Dictionary of Contemporary English* (3rd edition), *Oxford Advanced Learner’s Dictionary* (5th edition), and *Cambridge International Dictionary of English* (1st edition), all published in 1995. The vocabulary thus identified, which Stein calls the Common Core Vocabulary (CCV), has 2,139 lexical items. The majority of the list

are nouns (46.5%); other word classes are verbs (22.9%), adjectives (16.5%), and grammatical words (13.9%). The vocabulary is claimed to fulfill the following four criteria (Stein 2002b, 51–52):

- a. It is part of the standard vocabulary, which is “the most widely understood form”;
- b. In terms of style, it belongs to the unmarked vocabulary, which is widely known and causes “least offence”;
- c. The vocabulary is small, so that learners can grasp it within a short time. It is important for both the learner and the teacher that the vocabulary have a definite size;
- d. The vocabulary has to be expandable to minimize the learning effort. Expansion can be carried out by taking into consideration the meaning and form of a lexical unit.

Stein discusses at length criterion (d) above by presenting a great variety of strategies for vocabulary expansion. They involve the use of thematic domains, lexical fields, word families, correlative sets, word-formation, phonetic relations, among other strategies. This part of her book is an informative account of how to develop a more advanced vocabulary systematically on the basis of the core vocabulary. The author suggests that the vocabulary can be of great service to coursebook writers and teachers.

2.1.8 A New Academic Vocabulary List (2013)

In 2013 Dee Gardner and Mark Davies compiled The New Academic Vocabulary List (AVL) (Gardner and Davies 2013). The list was intended for academic purposes: it contained an academic core vocabulary required in English language education. It was based on an academic corpus of 120 million words, which was part of the 425-million-word COCA (Davies 2012). To obviate the methodological problems pertinent to the compilation of the AWL (see Section 2.1.6.), the researchers took into account lemmas, rather than word families, as a basic unit of the list. Furthermore, the list distinguished between different parts of speech. The AVL vocabulary was selected according to four criteria:

- a. Ratio, or relative frequency: the words (lemmas) must occur sufficiently more frequently (by at least 50%) in the academic corpus than in the non-academic part of COCA;
- b. Range: the words must occur frequently enough (20% or more of the expected frequency) in the majority of the academic disciplines (at least 7 out of 9);
- c. Dispersion: the words must occur uniformly across the disciplines;

- d. Discipline measure: the words cannot belong to a technical register, that is, they cannot occur “more than three times the expected frequency (per million words) in any of the nine disciplines” (Gardner and Davies 2013, 12).

The above statistical criteria allowed the researchers to identify those words that are typical of the entire academic register, and to exclude the words that are too technical or discipline-specific (Gardner and Davies 2013).

The researchers compared the AVL with the AWL in terms of coverage in major academic corpora. To make a valid comparison, they converted the AVL lemmas into word families and took into consideration only the top 570 word families, a number equal to the total of word families in the AWL. The conversion was carried out using over 20,000 word families compiled by Nation (Gardner and Davies 2013). The top 2,000 of the AVL word families are available at www.academicwords.info. The comparison revealed that the AVL word families cover twice as many running words in the COCA academic texts (13.8% vs. 7.2%) as the AWL word families (Gardner and Davies 2013). A similar coverage was reported in the academic texts of the BNC (13.7% vs. 6.9%). The above figures reflect the usefulness of the AVL for language pedagogy. According to the designers, the value of this list lies in the fact that its construction was based on a large contemporary corpus of unprecedented size, statistically viable measures for identifying the core words (frequency, range, dispersion, and discipline measure), and the use of lemmas.

2.1.9 A New General Service List (2013)

Sixty years after the publication of West's GSL, Browne, Culligan, and Phillips announced the compilation of a New General Service List (NGSL). The vocabulary was intended for second language learners of English. The compilers aimed to achieve a higher level of “the generalizability and validity” of the list by using an updated and largely expanded corpus compared to that used by West in the GSL (Browne 2013, 14). They used a 273-million-word subsection of the Cambridge English Corpus (CEC) (Browne 2013). The CEC is a 1.6-billion-word corpus of British and American English (Browne 2013), representing both written and spoken language. The words were drawn from the following registers: learner language, fiction, journals, magazines, non-fiction, radio, spoken language, documents, and television. Unlike the GSL, the UWL, and the AWL, the compilers counted lexemes, not word families. However, they adopted a modified definition of lexeme. Normally, the term is used to denote a lexical unit that comprises all its word forms which are of the same part of speech; for example, the verb lexeme *play* encompasses its various inflected forms: *plays*, *played*, and *playing*. Thus, part-of-speech homographs (e.g., *play* v. and *play* n.) are counted as separate

lexemes. Contrary to this approach, the compilers of NGSL treated part-of-speech homographs as separate lexemes. Therefore, under the lexeme *list* they grouped verb and noun word forms: *lists*, *listed*, *listing*, and *listings* (Browne 2013, 15).

A comparison of the NGSL with the GSL showed significant differences. The vocabulary of the NGSL made use of more word families (2,368) than the GSL (1,964), and provided greater coverage (90%) of the CEC than West's list (84%) (Browne 2014, 45–6). The increased coverage by the NGSL was not particularly surprising given the increase in the number of word families (by 368) compared to the GSL (see also Youngblood and Folse 2017). However, when one considers lexemes, the advantages of the list become more obvious, as the NGSL employed approximately 800 fewer lexemes than the GSL, and achieved a higher coverage (by 6.1%) of running words than West's list (Browne 2013).

As we have seen in this section, the interest in vocabulary control has not died out since the publication of the GSL in 1953. A number of new vocabulary lists have been created for educational purposes, using increasingly larger corpora and different methodologies. Some of the lists have built upon the earlier vocabularies, others (e.g., the Threshold Level English) have been compiled from scratch. However, the main principles underlying the design of the lists are easily traceable to West, Ogden, Palmer and other pioneers in the field.

2.2 The GSL against other word-lists and corpora

The GSL has attracted a great deal of attention from researchers, who have compared it with other vocabularies with respect to lexical content and the potential usefulness for contemporary students and teachers. Gilner (2011) provides an informative survey of studies evaluating the content of the GSL against other word-lists, which were conducted mainly by Paul Nation and his colleagues. One such study, carried out by Nation and Hwang (1995), shows a high overlap of the GSL with two other word-lists based on frequency and range: one retrieved from the Lancaster-Oslo-Bergen corpus (LOB) (Johansson 1978) and the other, from the Brown corpus (Francis and Kucěra 1964). Both corpora contained around 1 million running words. The word lists were compiled by extracting top-frequency words that occurred in at least 10 (out of 15) subsections in each corpus. The study demonstrated that the LOB list shares approximately 82.5% of its vocabulary with the other word-lists, the Brown list shares 65.1%, and the GSL, 70.5% (Gilner 2011, 73). From this last figure, it transpires that as many as 1,500 items (out of a total 2,147) in the GSL occur in the LOB and Brown lists. Gilner concludes that this finding is consistent with the fact that 1,500 words in the GSL are most frequent in English according to the Faucett-Maki list (1932).

The GSL has also been examined with regard to text coverage. In a study of the coverage of the LOB corpus, Nation and Hwang (1995) find that 2,147 word families of the GSL cover as much as 82.3% of the running words of the corpus. In turn, Hirsh and Nation (1992) report that the GSL covers approximately 90–92% of the running words in fiction, and Hirsh (1993) claims 95% coverage of this genre. As for non-fiction, Hwang (1989) reports the coverage of 75% by the GSL.

In another study, Sutarsyah, Nation, and Kennedy (1994) demonstrate that the GSL provides extensive coverage of two corpora: English for Academic Purposes (EAP) and English for Specific Purposes (ESP). The former corpus was composed of a collection of short academic texts from a variety of fields (medicine, mathematics, natural sciences, social sciences, law, education, among others), and the latter, a single textbook on Economics. Both corpora contained the same number of word tokens: 300,000. The aim of the study was to explore to what extent the vocabulary of general academic English (EAP) overlaps with the vocabulary of one specialized academic field (Sutarsyah et al. 1994). The results revealed little overlap between the corpora, suggesting that “EAP courses are of little value for learners with specific purposes” (34). Yet the researchers found that the GSL provided great coverage of both corpora: 78.4% of EAP, and 82.5% of ESP, and concluded that “to know most of the vocabulary in a specialized text a learner of English would need a vocabulary of at least 4,000 to 5,000 words.” (48). Because such texts vary in lexical content, it is impossible to predict exact words they contain. Therefore, the researchers suggested that the vocabulary absolutely needed by the university student should be the 2,000 word families of the GSL together with the 800 words of the UWL mentioned earlier.

In another research, Nation (2004) set out to examine the usefulness of high-frequency word lists developed from the BNC for syllabus design for EFL and ESL learners in primary and secondary schools. He compared three BNC-based lists, each containing 1,000 lemmas, with the vocabulary created through the amalgamation of the GSL and the AWL. With its largely non-academic vocabulary, the GSL was complementary to the more focused AWL. The latter contained a vocabulary that was considered important for senior high school and university students because the words it contained occurred frequently across a wide range of academic texts (Nation 2004). Words selected for inclusion in the BNC lists were lemmatized and rearranged according to frequency, range, and dispersion (i.e. evenness of distribution) as provided by Leech et al. (2001). Lemmas with the highest rates of the parameters had priority for inclusion in the lists. To perform fair comparison, lemmas were further rearranged into word families following Bauer and Nation (1993) (Nation 2004). The researcher found that despite its outdatedness, the vocabulary of the GSL, when combined with the AWL, covered most of the three BNC lists. With a total of nearly 2,000 word families in the

GSL, 97% of the first 1,000 word families were in the first two BNC lists. However, although the two sets of lists (the BNC lists and the GSL+AWL) contained much the same vocabulary, the vocabulary was not distributed identically in each set. In conclusion, Nation recommends that foreign learners in primary and secondary schools use materials based on an updated version of the GSL, while the AWL should be introduced at senior high schools and universities. In turn, materials based on the BNC lists are suitable for students the beginning of tertiary level.

In the same study, Nation compared the GSL, the AWL, and the BNC lists with respect to their coverage of four corpora. Each corpus represented a different register: written academic, technical (economics), spoken, and fiction, respectively. The study showed a remarkably high coverage by the GSL: 75.5% of the academic corpus, 82.5% of the corpus of economics texts, 89.6% of the spoken corpus, and 87.1% of the fiction corpus. The corresponding rates for the first two BNC lists were slightly higher, except for fiction: 83.9%, 89.8%, 91.1%, and 86.6%, respectively. Text coverage provided by the combined list of the GSL and the AWL resembled that of the BNC lists.

The above studies demonstrate two qualities of the GSL: high text coverage and high overlap with other word-lists. Thanks to these qualities, the GSL has enjoyed a long-established status in language pedagogy. It is remarkable that in spite of the fact that the GSL was compiled more than half a century ago, the list is similar in content to other word-lists (the LOB, the Brown, and the BNC lists). It should be noted, however, that the similarity is perceived in lexical form rather than meaning, as the latter is not so easily amenable to automatic comparison. The GSL provides comprehensive coverage of a variety of text types, ranging from 75% to 92% (see also Nation and Waring 1997). If supplemented by the AWL, the coverage is similar to that of the BNC lists. The above qualities arise from the fact that “the GSL is simply the result of it being, first and foremost, a frequency-range-based word-list” (Gilner 2011, 79). When word-lists are prepared according to the same criteria of frequency and range, the lists will be to a large extent similar irrespective of the source corpus (Gilner, 79). Of course, time span between the lists makes them to some extent different in lexical content. However, as Gilner concludes, although the language has changed over the period of the last century with respect to vocabulary use, the most frequent words are apparently least influenced by this change.

The qualities of the GSL presented in this section explain the long-lasting popularity of the list in language education. Towards the turn of the century, the GSL served as the basis for the *Longman Structural Readers* (Carter 2012, 198). It has been regarded as “an essential component of any initial English course” (Sutarsyah et al. 1994, 45). According to Nation (2011, 134), the list has proven far more useful than frequency-based lists made from corpora (e.g., Francis and Kučera 1982;

Leech et al. 2001). In the 1970s it served Longman lexicographers as the major source for the *LDOCE1* DV (see 3.2.2.2), and became a benchmark against which to evaluate other lists. We will see that a remarkably large part of the GSL remains valid today (see 4.3.4).

2.3 Controlled vocabularies for professional and human-computer communications

In the preceding section, we saw that language education provided a fertile ground for the development of controlled vocabularies. By controlling the vocabulary to which learners were exposed, texts had a very high chance of being comprehensible and less ambiguous. Language education is not the only area of application of controlled vocabularies. In fact, any user of a natural language may experience comprehension problems because of its ambiguity, imprecision, and complexities, which are inherent properties of such a language.

There are two major goals of creating controlled vocabularies: to improve readability and intelligibility of texts to language users, especially non-native ones, and to make texts easier for computers to process (Huijsen 1998; Wojcik 2006). According to the above goals, controlled vocabularies can be divided into human-oriented and machine-oriented (Nyberg et al. 2003). In what follows, I will discuss them in the context of, first, service manuals intended for human readers, and then texts to be submitted to machine translation (MT) systems.

2.3.1 Controlled English for writing service manuals

Restrictions imposed on a vocabulary and syntactic constructions are the basis for a controlled language. Ogden's Basic English is an early example of such a language, using a limited vocabulary of 850 words (Section 1.4). Basic English was created for educational purposes but it was not until the late 20th century that it inspired innovations in industry.

The Caterpillar Tractor Company was the first to create a controlled language for industrial purposes (Kamprath et al. 1998). Known as Caterpillar Fundamental English (CFE), the language was developed in the 1970s. The company is a manufacturer of heavy equipment and its components such as engines, hydraulic systems, and electrical parts. Their maintenance requires the production of multilingual technical documentation. In order to support international dissemination of the documentation, the company developed CFE. It was the language of service manuals, designed with non-English speakers in mind, who were responsible for reading and understanding the texts (Kamprath et al. 1998).

It was believed that the production of simplified CFE manuals would eliminate the need for their translation. Inspired by Ogden's Basic English, the company employed a vocabulary of the same size: around 850 items. The text of service manuals was flanked by numerous illustrations so as to help technicians with its interpretation (Kamprath et al. 1998). In 1982 the company abandoned CFE. One of the reasons was the problem of enforcing restrictions on the documents produced. Numerous English documents failed to comply with the CFE guidelines, even though they were marked as "CFE compliant" (Kamprath et al. 1998, 52). Another reason was that the restricted vocabulary was no longer sufficient for the rapidly expanding domains of high-pressure hydraulics and electronics. With the development of computer technology, the company decided to redefine the approach to the distribution of service manuals. In the 1990s, it designed Caterpillar Technical English for a new generation of controlled languages intended for use in machine translation.

Another branch of industry that benefitted from a controlled language was aerospace. In the late 1970s, the Association of European Airlines was concerned with the readability of aircraft maintenance manuals. The problem was that English was not the native language for many technicians who had to read the manuals. Thus, it was of paramount importance that the texts be written in simple English, as consequences of their misinterpretation would be disastrous and costly. There was a risk that any mistake in their comprehension might cause accidents. The Association entrusted researchers from the European Association of Aerospace Industries with the task of investigating controlled vocabularies. In collaboration with the Aerospace Industries Association of America, the researchers began a project which in 1986 led to the development of Simplified English Guide. It was a standard for writing manuals, which in 2005 came to be known as the ASD Simplified Technical English (STE), or ASD-STE100.

STE consisted of writing rules and a dictionary. The latter specified so-called approved words along with their approved forms, meanings, and examples of use (ASD-STE100 2017, i). The vocabulary consisted of about 850 general lexemes, which were "most frequently used in technical writing" (ASD-STE100 2017, 1-1-2). They were selected because they were "simple and easy to recognize" (ASD-STE100 2017, ii). They included lexical and function words, as well as combinations of words, for example: *a, about, find, pull, keep, approval, aid (n.), make sure, emergency, auxiliary (adj.)*.

Writers were allowed to use words, whenever possible, in just one meaning and one part of speech. For example, the approved meaning of the verb *to fall* was that of *to move down by the force of gravity*, and not *to decrease*. There were constraints on the selection and use of near-synonyms, for example *start* and not *begin, commence, initiate, or originate* (ASD-STE100 2017, ii). In order to ensure a

consistent use of such items throughout the text, STE suggested replacements for unapproved lexemes; for example, the verb *add* for the noun *addition*, *make sure* for *account for*, *assemble* for *build*, *but* for *on the contrary*, *transmit* for *carry*, and *layer (n)*. for *coat*. American English spelling was preferred to British English, with the forms permitted being those in *Merriam-Webster's Dictionary*. Word forms were kept under control; for example, the only acceptable forms of *adapt* were *adapt*, *adapts*, and *adapted*, whereas the participle form was not because it was liable to lead to confusion and misunderstanding (ASD-STE100 FAQ). Recently the use of affixes has become more restricted. In the 2017 version of STE, there are no affixes on the list, unlike the earlier ones, which included the prefixes *pre-* and *post-*.

In addition to the dictionary words, the writer was allowed to use “technical names” and “technical verbs” from outside the dictionary. These, however, could be used on the condition that they fell into one of pre-established categories. For example, from the category of “names of vehicles or machines and locations on them”, the writer could choose, among other terms, *aircraft*, *aircraft carrier*, *air-frame*, *airplane*, *bicycle*, *cabin*, *car*, and *cargo compartment* (ASD-STE100 2017, 1-1-5). The selection of the terms was further controlled by writing rules, for example:

- Use technical names that agree with approved nomenclature.
- When you must select a technical name, use one which is short and easy to understand.
- Do not use slang or jargon words as technical names.
- Do not use different names for the same item.
- Do not use technical verbs as nouns. (ASD-STE100 2017, 1-1-1)

Nominal compounds were allowable as long as they consisted of up to 3 nouns (e.g., *runway light connection*). This is because longer compounds could be confusing to non-native readers. When the writer could not avoid mentioning a complex nominal expression, they were instructed to use prepositional constructions (ASD-STE100 2017, 1-2-1); for example: *Calibration of the resistance of the runway light connection* instead of *Runway light connection resistance calibration*.

In order to verify the compliance with the standard, authors had checking tools at their disposal, such as the Boeing Simplified English Checker and MAXit Checker (Wojcik 2006). The tools were able to identify all the words that did not belong to the approved vocabulary, while suggesting alternatives. However, the detection of word meaning violation was more challenging (Wojcik 2006).

STE provided a model for other controlled languages. Those developed for industry include Ericsson English, General Motor's Controlled Automotive

Service Language (CASL), IBM's Easy English, Kodak's International Service Language, Océ's Controlled English, Sun Microsystems' Sun Controlled English, and Xerox Multilingual Customized English (Muegge 2008). Unlike STE, most of the languages are proprietary to specific companies, which is indicated by their names.

Another controlled language is Perkins Approved Clear English (PACE). PACE was a simplified English introduced in 1980 at Perkins Engines Ltd, an American manufacturer of engines and machines. According to Pym (1990, 80–82), the then manager, PACE was intended for non-native engineers, especially those who had to read the English documentation for the company products. Originally, the simplified language was used to facilitate human comprehension of texts, but it soon became clear that it lent itself to controlling input texts for machine translation. Thus, in 1985 the company tested the quality of the translation of PACE-controlled documents, produced by a MT system (Weidner's MicroCat), and concluded that the controlled language was the “ideal input” (cited in Crabbe 2017, 32).

PACE was composed of a controlled vocabulary of approximately 2,500 words and a set of simplified rules for writers (Nyberg et al. 2003; Pym 1990). Below are examples of rules:

- Keep sentences short
 - Omit redundant words
 - Avoid elliptical constructions
 - Adhere to the PACE dictionary
 - Avoid strings of nouns
 - Do not use ‘-ing’ unless the word appears thus in the PACE dictionary
- (Pym 1990, 85–86)

A fundamental rule for the PACE vocabulary is “one word for one meaning” (Pym 1990, 85). The system comes with a dictionary that provides definitions for selected words, “where they are necessary”. Thus, to avoid misunderstanding, the system provides definitions for “common words” and “engine parts with common names” (e.g., **part**, n. “A service part”, **pass** vb “To proceed”, **passage** “A drilling along which a fluid moves”) (Pym 1990, 85). There are no definitions for technical terms (e.g., *parent bore*), as they are assumed to be familiar to the staff. Noun clusters are allowable, but they should “consist of no more than two or, at the most three words” (Pym, 86). Present participles are not permitted because of their ambiguity. According to Pym, post-editing of the translation of PACE-controlled texts is significantly faster than of the texts written in a conventional way (Pym, 86).

Controlled languages like STE and PACE were originally devised as standards for writing manuals, and only later became useful for controlling input texts for machine translation. The latter use of controlled languages is the subject of the following section.

2.3.2 Text production for machine translation

Controlled vocabulary is noteworthy for its place in the history of machine translation. Until the development of neural machine translation (NMT), which is state-of-the-art translation technology, many MT systems employed controlled vocabularies to improve the quality of their translations.

The idea that a natural language can be translated automatically with no human intervention has been appealing for ages. However, it was not until the 20th century that researchers offered the first practical suggestions for a machine translation system. The first public demonstration of such a system was held in New York in 1954 (Hutchins 1995). The experiment involved the use of a computer program which produced a word-for-word translation of 49 Russian sentences into English. The translation was performed with a limited vocabulary of 250 words and a few syntactic rules. Although the experiment attracted publicity and inspired further research in the US and abroad, it was less worthy of attention from a scientific point of view because the test material was preconceived to eschew ambiguity problems. The sentences displayed fairly simple structures, with verbs used only in the third person singular or plural, and with no negative particles (Zarechnak 2000). There were no compound or interrogative sentences. Lexical and syntactic restrictions appeared to be necessary for any translation system to work successfully.

By that time, it had become clear that extensive research was needed to produce a fully automated translation system. Researchers were concerned with the poor quality of word-for-word translations. Suggestions were made to restrict the language of the source text and to develop sublanguage systems (Hutchins 1995). For example, in 1952 Stuart Dodd proposed rules for controlling the language of texts submitted to MT. In his “Model English”, each word was supposed to have a fixed position in a sentence, and be used in just one meaning, and one grammatical, phonetic as well as spelling form (Dodd 1955).

By the mid-1960s, many countries around the world launched MT projects, using either the above direct translation approach or developing other rule-based models (Hutchins 1995). They relied on the formulation of rules for different levels of language analysis, from morphology, and syntax, to semantics and phonology. They employed bilingual dictionaries for selecting equivalents. Research on MT translation progressed at a steady pace, sometimes by trial and error, but the quality of translations did not come close to that of the human translation. For example, a system designed for the US Air Force around 1960 generated output that was sometimes barely comprehensible, yet sufficient for scientists to meet their basic information needs (Hutchins 1995).

Nevertheless, it was observed that while MT systems were unable to cope with complexities of the general language, they produced quite accurate results for specific language domains. In the 1970s a translator in the Canadian Meteorological Center remarked that weather forecasting lends itself to machine translation because it is repetitive in structure and limited in vocabulary (Thouin 1982). In 1976 a research team at the University of Montreal developed *Météo*, a system for the automatic translation of weather forecasts between English and French. Initially, it was used for translating messages from two regional offices in Halifax and Toronto, but soon it became an integral part of the Canadian network for the transmission of meteorological reports (Thouin 1982). The system was in constant use every day and night. Designed to process forecast bulletins, it was capable of identifying a few thousand words and a small number of sentence patterns. It was able to recognize the sections of a bulletin (i.e. headings, title, list of regions, and forecast), and to distinguish between messages intended for translation (weather forecasts for various audiences) from those it was unable to process (e.g., synopses) (Thouin 1982). The latter were automatically rejected by the system and submitted to human translators. According to Thouin (1982, 43), a large body of the rejected texts did not conform to “writing and format standards for weather bulletins”. The output of *Météo* was reasonably good, and around 80% of the words submitted to the system were translated every day without human correction. The system was appreciated for relieving translators of repetitive work and ensuring consistency in the use of terminology, while increasing the speed of translation. The success of the system depended largely on lexical and syntactic restrictions of meteorological reports. *Météo* was in use until 2001.

Another MT system worth mentioning is Systran. Although designed to translate the general language, in practice it worked with specialist subject domains (Hutchins 1995). Designed in 1968, it was installed in 1970 at the US Air Force to translate technical and scientific documents from Russian into English, and in 1976 at the Commission of the European Communities for English-French translation of legislation. Since then Systran has undergone numerous modifications and improvements, which extended its applicability to the translation of English into a large number of languages. Although traditionally rule-based, it has recently employed a hybrid approach to MT, combining rule-based and statistical methods. Large companies such as General Motors of Canada, the Xerox Corporation, and Google (Google Translate until 2007) have used the system. By restricting the vocabulary and syntax of the source language, Xerox has been able to practically eliminate post-editing of translations (Hutchins 1995).

The finding that controlled input significantly improves the quality of MT led to the development of KANT (Sin-Wai 2015; Allen 1995). KANT was a knowledge-based system, which employed a metalanguage in the form of networks of propositions (Qun and Xiaojun 2015). Launched in 1989 at the Center for Machine Translation of Carnegie Mellon University, the system employed a strictly controlled vocabulary and grammar for a specific subject domain. Meanings were encoded for the domain in advance, and during translation served as points of access to the knowledge base of that domain (Mitamura 1999). The system retrieved domain-specific terms from a corpus, and encoded, whenever possible, a single meaning to each term. This reduced semantic ambiguity and increased translation accuracy. The 1992 prototype of the system made use of several hundred technical terms and 14,000 meanings for general words.

When a term was used in more than one meaning in the domain, it was possible to encode each meaning in separate entries. The system performed analysis of a source text to check whether its vocabulary conformed to the pre-established constraints, and in the case of lexical ambiguity, allowed the author to solve the problem by selecting an alternative word (Mitamura 1999). The author intervened in the text in the interactive mode. However, even without human disambiguation and post-editing, the KANT prototype from 1992 produced high-quality translations of technical electronics manuals from English into German, French, and Japanese (Nyberg and Mitamura 1992). This was largely due to restrictions on input texts.

The system imposed constraints on the use of pronouns, conjunctions, and present as well as past participles in a source text, for they “increase the potential ambiguity in syntactic analysis” (Mitamura 1999, 47). Present participles with no explicit subject, as in *When starting the engine*, were not allowable (Mitamura 1999, 47). Writers were advised to avoid reduced relative clauses (e.g., *the pumps mounted to the pump drive*) and to use instead that-clauses with the explicit pronoun *that* (*the pumps that are mounted to the pump drive*) (Mitamura, 47). The use of acronyms was kept under control; for example *OF* was not allowable because the computer could interpret it as the preposition *of*. Writers were asked to use only one spelling variant consistently throughout the text, and to avoid phrasal verbs, as they might be misinterpreted by the computer as combinations of verbs and prepositions.

KANT paved the way for the development of CATALYST. Designed by the Carnegie Mellon specialists, the system is one of the largest knowledge-based interlingua MT systems used in translation of user and maintenance manuals. With controlled language input, it generates high-quality translation needing little post-editing.

The invention of neural MT in recent years is a major step forward in translation technology. NMT systems incorporate neural networks which operate within

encoder-decoder framework. Roughly speaking, the encoder network reads a source sentence and renders it in a set of numbers; then the decoder network decodes the numbers and generates a target sentence (Cheng 2019). NMT systems often deliver satisfactory translations needing little post-editing, but the risk of errors remains relatively high, and varies from one language to another (Benjamin 2019; Khoong et al. 2019). The systems do not need controlled vocabularies, and are capable of high-quality translation without such restrictions. However, there are studies demonstrating that controlled input results in an improved accuracy of translation of morphologically rich languages (e.g., Duygu et al. 2017).

The history of machine translation shows that many systems have used a controlled vocabulary to improve translation quality. Along with the restriction of source texts to a language domain, controlled vocabularies made translation more efficient and effective. They significantly reduced the time and cost of translation, while improving readability, comprehensibility, and reusability of texts (Sin-Wai 2015).

2.4 Identifying a core vocabulary

So far we have been looking at various controlled vocabularies compiled for pedagogical and communicative purposes. Underlying these lists is the idea of essential, basic, or core vocabulary. This idea is fundamental to the composition of, *inter alia*, Eldridge's *Six Thousand Common English Words*, Thorndike's *Word Book*, Ogden's *Basic English*, and West's *General Service List* (Stubbs 1986). Although the lists vary in lexical content, the concept of coreness is intuitively recognizable.

Minkova and Stockwell (2006, 463) describe a core vocabulary as one which is crucial for communication, "without which sentence composition [...] would be unthinkable". In turn, Bell (2013) argues that it is well known to all adult native speakers. In the current literature, writers prefer the term *core vocabulary* to its earlier variants, such as "essential", "nuclear", "procedural", to mention but a few (Stubbs 1986; Bell 2013). However, the concept of core vocabulary is by no means novel.

The question of what constitutes a core vocabulary has been much alive among lexicographers and educational linguists for over a century. As we have seen in Chapter 1, the central concern of language scholars and teachers in the first half of the 20th century was how to establish a basic or minimum vocabulary that would help improve the learner's competence in reception and production. The concept of core vocabulary appears in 'General Explanations' which opens the first volume of the *OED*. The founding editor of the dictionary, James Murray explains that "the English vocabulary contains a nucleus or central mass of many thousand words [...] they are the common words of the language" (Murray 1888, xvii).

Murray's observation is followed by the famous compass-like drawing illustrating the editor's vision of the lexical composition of the English vocabulary (Figure 1). The diagram displays common words in the very center; colloquial and literary somewhat further away from it; and scientific, foreign, dialectal, slang, and technical pointing out to the periphery (see also Curzan 2000; Aarts and MacMahon 2006). The diagram unequivocally suggests that for Murray the relationship between the nucleus and common words was straightforward.



Figure 1. Murray's vision of the English vocabulary (from Murray 1888, xvii).

As Carter observes, there is no one core vocabulary, but many, depending on the circumstances of communication (2012, 47). It is the communication type that determines the criteria for vocabulary restrictions. Only within the pedagogical field, word-lists are created for a variety of purposes: preparing teaching materials for foreign language learners, checking their lexical appropriateness for different groups of learners, preparing vocabulary tests, setting up an expected level of linguistic competence for native English-speaking school students (Stubbs 1986). For example, the GSL served not only as a basis for graded readers, coursebooks, and dictionary definitions, but also as a medium of instruction in the classroom. It was developed to facilitate receptive and productive needs of learners.

Stubbs (1986) and Carter (2012, 1986) propose more precise guidelines for establishing a core vocabulary for learning and teaching. Stubbs offers 12 criteria for determining what he calls the nuclear vocabulary, and Carter suggests 10 tests for coreness. Taken together, all these criteria and tests fall into two groups: the first one defines coreness in terms of syntactic and semantic relations between words, and the other, in terms of various conceptions of word neutrality. As Carter (2012, 47) explains, the former group comprises words that are “more tightly integrated than others into the language system”, and the latter group consists of words that are “unmarked and non-expressive” in discourse. Although both researchers admit that word frequency is central to identification of core words, they do not treat it independently of the other criteria. According to Stubbs (1986), frequency and other quantitative parameters are a consequence, not a cause, of word coreness.

Because the criteria proposed by Stubbs and Carter overlap to a large extent, they were summarized by Bell (2013) with a unified terminology. The presentation below employs Bell's terminology with some modifications. Thus, the first group mentioned above comprises criteria that refer to syntactic and semantic relations:

- a. **Superordinateness.** Core vocabulary items are more likely to be hypernyms (also called superordinates) than hyponyms. For example, the core words *flower*, *house*, *animal* tend to be classified as hypernyms, whereas *tulip*, *chalet*, *tiger* are hyponyms. While hypernyms can be found at both higher and lower levels of the semantic hierarchy (*tulip* is in turn a hypernym of *dwarf tulip*, *lady tulip*, *cottage tulip*), finding a hypernym among core words is much easier than among non-core ones. The lower-category words such as *tulip* and *rose* are less likely candidates for being core words because finding their hyponyms is not easy for the average English speaker.
- b. **Semantic productivity.** Core words tend to have more meanings than non-core words. The rich polysemy of the former is reflected in the length of dictionary entries in general-purpose dictionaries (Bell 2013). It is easy to observe that the entries for such words record numerous senses.
- c. **Formal productivity.** Just as core vocabulary items are semantically productive, so they also frequently participate in the creation of set expressions such as compounds, idioms, and phrasal verbs. Again, this is reflected in large-size dictionaries, which provide a number of set expressions in the entries for core words.
- d. **Potential for substitution.** Core words, such as *run*, can substitute for non-core ones, such as *sprint*, *trot*, *race*, *gallop*, *hurry*, but the reverse is not necessarily possible without a context (Bell 2013). The potential of words for substitution can be checked by writing definitions of words belonging to the same lexical set⁴ and determining which of these words are indispensable and which can be omitted. For example, Carter reports (2012) on an experiment, in which he tested the substitutional power of words by asking a group of informants to produce definitions of words belonging to a lexical set. One of the sets was the following: *guffaw*, *chuckle*, *giggle*, *laugh*, *jeer*, and *snigger*. It turned out that the majority of the subjects (80%) defined the non-core verbs with *laugh*, the verb which emerged as the most core word in the set. The respondents' definitions displayed a syntactic pattern which consisted of a core verb followed by an adverb or adverbial phrase, for example, **chuckle** "laugh quietly" and

4. A lexical set is a group of lexemes which share a common semantic feature, for example, ways of walking, colors, emotions, vehicles, and time.

guffaw “laugh in a loud and rude manner” (Carter, 50). Defining the other way around, for example, **laugh** in terms of *chuckle* or *guffaw*, would probably be possible,⁵ but this strategy was clearly not what the respondents preferred. Other lexical sets presented to the respondents are shown below, with the core words indicated in bold:

perambulate, stroll, saunter, **walk**, hike, march
 podgy, corpulent, stout, **fat**, overweight, plump, obese
 weedy, emaciated, skinny, lean, **thin**, slim, slender
 adobe, **house**, domicile, residence, dwelling (Carter 2012, 50)

The above test was essentially based on a paraphrase within simple words, and as such is familiar to EFL lexicographers. West used this strategy extensively in developing his vocabulary list.

- e. **Collocability.** This criterion is based on a hypothesis that core words have a higher collocability than non-core ones. Collocability is the potential of words to collocate, that is, to co-occur with a particular range of words (Sinclair 1991; Sterkenburg 2003). Words differ with regard to collocational range; for example, it is much easier to find in the BNC collocates for the adjective *false* (e.g., *statement*, *impression*, *teeth*, *pretences*, etc.) than for its partial synonyms *fake* (e.g., *alarm*, *notes*) and *bogus* (e.g., *claims*) (Kamiński 2016). Thus, out of these near-synonyms, *false* emerges as the core word because it attracts a wider range of collocates than the other words.

It can be argued, however, that collocational range of words is a reflection of their polysemous nature. Thus, the argument goes, the validity of a comparison of collocational ranges of two words is diminished by the fact that they collocate in different meanings. For example, *shiny* enters into collocations with *nose*, *coin*, and *car* because its meaning is different from that of *bright*, *radiant*, and *gaudy* (Carter 2012, 52). However, even when we disregard the problem of non-corresponding meanings, collocability is in a direct relationship with the aforementioned semantic productivity.

-
5. Defining a hypernym in terms of a hyponym is theoretically possible, for example by using the latter and negating the differentia that distinguishes it from the former. This approach has no place in general-purpose dictionaries, which are user-oriented. However, while in analytical definitions, hyponyms do not serve as *genus terms*, they can be used as *differentia specifica*. In particular, they can be used as extensional elements identifying typical examples of the category being defined and examples that do not fall into this category (Geeraerts 2003), for example, **animal** “a living creature such as a dog or cat, that is not an insect, plant, bird, fish, or person” (LDOCE4).

There is evidence that the more senses a word has, the broader the collocational range (Baker 1992). For example, the BNC provides evidence that the adjective *false* collocates with *statement* in the sense of “being completely untrue”; with *teeth* in the sense of “artificial”; with *impression* in the sense of “wrong”; and with *pretences* (in the expression *under false pretences*), in the sense of “deceiving people” (Kamiński 2016). The number of senses of *false* is significantly larger than those of *fake* or *bogus*. This difference is reflected in the number of senses covered by dictionaries, for example *LDOCE4* records 8 senses of *false*, 2 senses of *fake*, and 1 sense of *bogus*. Given the fact that the number of senses correlates with the collocational range, and in turn with the level of semantic generality, we can conclude that *false* is a better candidate for a core word than the other adjectives.

- f. **Antonymy.** Core words tend to have easily predictable antonyms (e.g., *fat* – *thin*), whereas it is more difficult to specify exact antonyms for non-core words. Thus, finding antonyms for *corpulent*, *obese*, *stout*, and *plum*, which are located at different points of the semantic cline, is more challenging than for *fat* (Carter 2012, 50; Bell 2013).

Underlying the second group of criteria for coreness is the concept of neutrality. The criteria below represent different types of neutrality: affective, cultural, contextual, and register (Bell 2013):

- g. **Affective neutrality.** Core words do not evoke affective connotations such as those triggered by feelings and attitude. In this regard, *old* is neutral, but *ancient* is not (Bell 2013).
- h. **Cultural neutrality.** Core words are not marked by cultural associations, whereas non-core ones carry cultural traits and tend to be borrowed from one language to another. Thus, as Stubbs (1986) observes, certain semantic categories such as basic bodily functions and activities (e.g., *sleep*, *eat*, *go*, *be*, *see*), size and shape (e.g., *big*, *round*), and natural phenomena (e.g., *sun*, *moon*, *earth*) are culture-free and resistant to interlingual borrowing. In addition, the stable core of the vocabulary of a language covers grammatical words, kinship terms, names of body parts, and frequent color terms (Trask 1996). On the other hand, certain cooking and furniture terms of French origin, such as *pouf* and *chaise-longue*, are poor candidates for the core vocabulary (Carter 2012, 54).
- i. **Contextual neutrality.** Core vocabulary items do not reveal the field of discourse from which they originate. Thus, while everyday words such as *brain*, *shin*, *bone*, and *stomach* do not evoke specialist contexts of use, the respective medical variants *cerebellum*, *tibia*, and *epidermis* are restricted to medical discourse (Stubbs 1986).

- j. **Register neutrality.** Core words are not associated with a particular register; they are neither formal nor colloquial or slang. Thus, core words *help* and *drunk* are used in neutral communicative situations, while *give a hand* is colloquial and *intoxicated* is formal (Stubbs 1986).

The above criteria are well grounded in linguistics. The criteria a–f are based on the paradigmatic and syntagmatic properties of words: they identify the position of a word in relation to other words. These criteria find support in structural semantics and its various sub-fields and theories: semantic field theory (e.g., Trier 1931), relational semantics (e.g., Lyons 1969; Lyons 1977), and componential analysis (Pottier 1964; Coseriu 1962; Nida 1975). What is common to the above approaches is that language is considered as a system governed by principles, with the linguistic sign relating to other signs in a principled way (e.g., synonymy, antonymy, hyponymy etc.). As for the criteria g–j, they refer to non-denotational aspects of meaning, which have been extensively discussed in the context of semantic change by historical philologists (e.g., Sperber 1914; Van Ginneken 1911; Shreuder 1929), and by more recent linguists identifying evoked and expressive types of meaning through a comparison of partial synonyms (Lyons 1977; Cruse 1986). The emotive meaning has received particular attention in the psycholinguistic studies of affective values of words (Osgood et al. 1957; Snider and Osgood 1969).

The guidelines proposed by Stubbs and Carter can be regarded as a linguistically-oriented elaboration of the principles of vocabulary control by the pioneers in the field. West employed the criteria in his work as a writer and lexicographer. He, as it were, tested the substitutional power of his controlled vocabulary by paraphrasing complex lexical units, replacing semantically narrow words with general ones, and preferring neutral words to those that were stylistically marked. Whenever possible, he avoided words restricted by discourse type. The guidelines lend themselves to being used by lexicographers in writing definitions. However, because the criteria are purely linguistic, they cannot be accepted for teaching purposes without reservations. This fact seems to undermine their application in definitions for learners.

One limitation of Stubbs' and Carter's criteria is that they completely disregard the learner's perspective. As Carter himself admits, they do not answer the question of which core words are easier to recognize or memorize for learners. The researcher calls for further investigation of learnability and teachability of a core vocabulary, as well as its memorization and lexical recall. He observes that highly polysemous words are most problematic because it is not exactly clear which meanings should be taught first (Carter 2012, 225). This question was also raised by Palmer and other pioneers of the vocabulary control movement (see 1.3.1. and 1.3.2.). A later study by Bauer and Nation (1993) continues this line of

investigation, and the studies by Webb and Nation (2008) apply predictability and other criteria to estimate the difficulty of reading texts.

A related problem is that the criteria for coreness presented in this section downplay the problem of multiple meanings. Although they do recognize that extensive polysemy is a property of core words (the principle of semantic productivity), they do not specify which meanings are core. To identify such meanings, one can adopt linguistic and lexicographic perspectives.

From the point of view of cognitive semantics, a core meaning of a word is psychologically salient and intuitively obvious. It is the one that is activated immediately in the native speaker's mind when they hear the word, and is usually the one they learnt first as children (Burkhanov 1998). In order to gain a clearer understanding of what a core meaning is and how it relates to other meanings of a word, one may look at the model of lexical category offered by prototype theory. According to this model, the semantic structure of a lexical category is flexible and fuzzy. Meanings display a radial structure, with one or more prototypical cores, from which other related meanings extend (Lewandowska-Tomaszczyk 2007: 145). They show degrees of typicality, as some meanings constitute better examples of the category than others. This model is particularly useful for lexicographers, whose task is to discriminate between various meanings of a word.

Indeed, prototype theory has attracted much attention from dictionary makers and theoreticians (Geeraerts 2006a; Hanks 1994; Rey 1990; Swanepoel 1994; Meer 1999 and 2000; Zgusta 2006). Geeraerts (2006a, 363) believes that the theory provides "an exciting perspective for the further development of lexicography", and Zgusta (2006, 115) remarks that it "promises to be of great usefulness to lexicography". In the compilation of *New Oxford Dictionary of English* (NODE1998), the editor declares to have taken full advantage of developments in cognitive linguistics when arranging entries and defining meanings (Preface). As Hanks, the editor, explains, "[e]ach entry has at least one core meaning, to which a number of subsenses, logically connected to it, may be attached" (Preface). Hanks goes on to explain that "[c]ore meanings represent typical, central uses of the word in question in modern standard English [...] The core meaning is the one accepted by native speakers as the most literal and central in ordinary modern usage" (Preface). This presentation of senses is supposed to reflect, at least in part, the prototypical organization of cognitive categories.

NODE aside, the history of lexicography shows that dictionary makers had already developed ways of discriminating between various meanings of words. Historical dictionaries, such as the *OED* (1888–1928), are a case in point. The methodology of discovering the history of words underpins the presentation of senses in an entry. Historical lexicographers peruse corpus evidence in order to determine one or more prototypical core senses and then possible paths

of development of other senses and sub-senses (Kay 2000). The literary record serves as a basis for mapping senses in a dictionary. When there is no sufficient evidence or there are gaps in the record, lexicographers infer the actual sequence using a logical principle; for example, by placing spatial meanings before temporal, concrete before abstract, general before specific, etc. (Fraser 2008). To reflect the chronological and multi-layered development of senses, the *OED* indicates senses with labels, such as capital letters, Roman and Arabic numerals, as well as small letters. Such an organization exposes original or historically earliest meanings, which come first in an entry. The identification of etymological meanings may have a pedagogical value, as they often contribute to the memorization and understanding of more recent related meanings (Ilson 1983). However, etymology may be misleading, as these meanings do not always coincide with those that are psychologically salient. For example, the origin of the English word *camera* can be traced back to the Latin *camera* “vaulted roof or chamber”. This meaning can be considered as the original one, but given the contemporary use of the word, it can hardly be treated as the primary one.

A learner using learners’ dictionaries may have an impression that the sense that comes first in an entry is the core one. This is not always the case, as learners’ dictionaries give priority to the most common meanings by putting them first in the entry. The underlying assumption is that the prominent position of this meaning will be convenient for readers of contemporary texts, provided of course that they start perusing an entry from its beginning. This approach is also pragmatically sound, as learners are most likely to look up the most frequent meanings. However, meaning frequency is not always a reliable guide to their centrality, because the most frequent meanings are not always central. It is not uncommon to find entries in which a metaphorical sense comes first, before a literal one. For example, *LDOCE3* begins the entry for *hackles* with the metaphorical sense of the idiom *sb’s hackles rise*, because it is most frequent, and continues with the literary sense of *hackles* as the second one (Stein 2002a, 117–8). Such a presentation of senses is counterintuitive because it destroys the logical flow of meaning (Moon 1987). The identification of the most frequent meanings has become possible through the availability of large corpora and sophisticated computer software. However, the identification is not without problems, especially in the case of highly polysemous words. While it is easy to compute the total number of occurrences of a lexeme, the automatic discrimination between its various meanings is challenging.

Word sense disambiguation (WSD) has been of great interest to computational linguists and computer scientists in the field of natural language processing (NLP). However, their approaches to WSD are not always sensitive enough to detect subtle differences in meaning. One of the most intractable problems with WSD is that “there are no decisive ways of identifying where one sense of a word

ends and the next begins” (Kilgariff 2006, 29). A number of NLP applications rely on a pre-established inventory of word meanings (Preiss and Stevenson 2004). One such a lexical database is WordNet (Fellbaum 1998), in which words are organized into synsets, that is, sets of synonymous words. The organization of WordNet into synsets makes the database particularly attractive to NLP developers, who treat it as a benchmark for identifying word meanings in texts. However, the indiscriminate equating of synsets with word meanings raises doubts about the soundness of this approach. As Hanks and Pustejovsky (2005, 65) explain, “Closer inspection shows that many WordNet’s senses are indistinguishable from one another by any criterion – syntactic, syntagmatic, or semantic – other than the fact that they happen to have been placed in different synsets”. Further problems emerge in the light of prototype theory; for example, given the fact that meaning boundaries are not always perceptible, there may be no corresponding senses in WordNet. The division of meaning into discrete senses raises doubts as to the treatment of boundary cases. Clearly, in this approach the quality of a source lexical database determines, but does not guarantee, the success of WSD.

Other approaches to WSD take advantage of software retrieving contextual clues, such as lexical collocation, grammar, and syntax, from a corpus (Kilgariff and Rosenzweig 1999). Sketch Engine is a corpus tool used by lexicographers to provide them with a snapshot of a word’s collocations and grammatical relations on the basis of a large body of texts (Kilgariff and Rundell 2002). Although extremely useful, such information is only a starting point for sense discrimination, the task which needs to be completed by a lexicographer. To lighten the lexicographer’s task of discriminating between meanings, Atkins et al. (2003) suggest an approach that draws on Fillmore’s theory of frame semantics (1995). Atkins et al. analyze the semantic content of *argue* and *argument* by investigating the semantic neighbours of the words. The researchers offer principles for WSD in the hope of making “it easier for computers to handle the drudgery involved” in dictionary making (Atkins et al. 2003, 280). The researchers believe that some stages in the analysis of meaning, which are normally left to a human being, are amenable to automatic processing. Another researcher drawing on frame semantics is Hanks (2019). In his *Pattern Dictionary of English Verbs* (2019) he distinguishes different meanings of verbs by classifying their collocates. Using an inventory of 230 basic semantic types, such as PERSON, ENTITY, EVENT, and STATE, Hanks assigns them to semantic roles depending on context. For example, “the old man”, which represents the semantic type *person*, may act as a *doctor* in the context of treating patients, and as a *patient* in the context of being treated by a doctor (Hanks 2015, 93). Such a database can be used by the computer in the classification of verbs according to meanings associated with prototypical sentence patterns. Research into WSD is promising but, as yet, the task remains an intractable problem. WSD

is not sensitive enough to ensure the marking of all words' meanings. In particular, metaphorically extended meanings and ad hoc creations are difficult to identify automatically.

Thus, a core form seems to be easier to identify than a core meaning. Although the concept of core meaning is reasonably tangible and easy to grasp, dictionaries do not always assign a central status to such meanings. The task of discriminating between core and non-core meanings is not always straightforward. It is even more challenging for computers, and much of the task is left to a human being.

2.5 Wierzbicka's Natural Semantic Metalanguage and relevant projects

The search for core concepts lies at the root of Wierzbicka's research into semantic description. In the late 1960s, Wierzbicka set out to define words within a restricted set of universal concepts. Her approach to the description of meaning, which is known as Natural Semantic Metalanguage (NSM), has been developed in numerous works (Wierzbicka 1972, 1985, 1996, 2003, 2014; Goddard and Wierzbicka 2002, 2014). NSM has stimulated interest of researchers in the fields of lexicography and cross-linguistics communication. The researchers have launched projects leading to the creation of, *inter alia*, Minimal English and a primitive-based dictionary. They will be discussed in this section.

2.5.1 Natural Semantic Metalanguage

Inspired by Leibniz's philosophy of language, Wierzbicka maintains that all concepts can be defined in terms of the most basic meanings, which are not definable themselves. These irreducible meanings are referred to as semantic primitives. NSM addresses the fundamental principle of defining, whereby a complex concept should always be defined in terms of simpler concepts, not the other way around (Wierzbicka 1985). This principle, expressed by Aristotle three centuries BC, underlies Wierzbicka's conceptual analysis. Like Aristotle's *priora* and Leibniz's *simples*, semantic primitives cannot be decomposed themselves but are used to articulate all complex meanings. A word meaning is explicated with primitives in a process called reductive paraphrase; for example:

sun

something

people can often see this something in the sky

when this something is in the sky

people can see other things because of this

when this something is in the sky

people often feel something because of this

(Wierzbicka 1996, 220)

According to Wierzbicka (1972, 1996), semantic primitives represent universal concepts, which have been lexicalized in all languages. Wierzbicka and her colleagues have identified an inventory of primitives empirically, through observation of a number of languages from typologically different groupings (Wierzbicka 1996; Goddard and Wierzbicka 2002). In 1972 Wierzbicka proposed 13 primitives, but over the decades the list has been revised and expanded, reaching 63 in 2010 (Goddard 2010), and 65 in 2014 (Goddard and Wierzbicka 2014). The list includes frequent words, both function and content ones, categorized into several classes:

substantives: I, YOU, SOMEONE/PERSON, SOMETHING, THING, PEOPLE, BODY
 relational substantives: KIND, PART
 determiners: THIS, THE SAME, OTHER/ELSE
 quantifiers: ONE, TWO, MUCH/MANY, SOME, ALL
 evaluators: GOOD, BAD
 descriptors: BIG, SMALL
 mental/experiential predicates: THINK, KNOW, WANT, FEEL, SEE, HEAR
 speech: SAY, WORDS, TRUE
 actions, events, movement, contact: DO, HAPPEN, MOVE, TOUCH
 location, existence, possession, specification: BE (SOMEWHERE), THERE IS/
 EXIST, HAVE, BE (SOMEONE/SOMETHING)
 life and death: LIVE, DIE
 time: WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME,
 MOMENT
 space: WHERE/PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE
 logical concepts: NOT, MAYBE, CAN, BECAUSE, IF
 augmentor, intensifier: VERY, MORE
 similarity: LIKE (Geeraerts 2010, 128–9)

From a lexicographic perspective, one cannot ignore advantages of this approach to meaning description. For one thing, the approach allows the lexicographer to overcome circularity, which permeates dictionary definitions; for another, it enhances the explanatory value of definitions (Geeraerts 2010). As Wierzbicka (1985, 49) maintains, the meaning of words, even of frequent ones, can be elucidated in basic everyday words, such as *glass* or *around*, without recourse to unnecessary technical jargon, such as *vitreous* or *cross-section*. Such definitions will be easily understood by the general reader.

From a historical point of view, Wierzbicka's quest for primitives is by no means original, which she herself admits. This line of research was pursued by Gottfried Leibniz, a philosopher of the 17th century. Leibniz set himself a goal of discovering a limited number of fundamental concepts, which could not be decomposed further but could be used as building blocks to form all possible complex concepts in a language (Taylor 1995). He believed that fundamental concepts would enable him to capture the "compositional nature of human thinking" (Maat

2004, 295). However, except for several incomplete drafts, he did not offer concrete solutions in this regard, nor was he able to specify the number or type of such concepts. As Wierzbicka points out, Leibniz “never proposed anything like a complete list of hypothetical primitives” (1996, 13). In more recent times, the Leibnizian interest in semantic primitives reemerged in the works by a Polish linguist Andrzej Bogusławski (1966, 1970), who, like Wierzbicka, represents the Polish Semantic School (De Schryver 2008). In the 1960s, Bogusławski maintained that researchers can approach the universal units of thought in a novel way by drawing on both empirical and theoretical achievements of modern linguistics. As Wierzbicka reports,

[t]he ‘golden dream’ of the seventeenth-century thinkers, which couldn’t be realized within the framework of philosophy and was therefore generally abandoned as a utopia, could be realized, Bogusławski maintained, if it was approached from a linguistic rather than from a purely philosophical point of view.

(Wierzbicka 1996, 13)

Wierzbicka argues that her definitions are not only accurate but also intelligible to the reader. As she maintains,

It is also my objective to show that the meaning of everyday words can not only be explicated in a theoretically justifiable way (without a guilty conscience) but also that it can be stated in simple and clear words, without any technical jargon or esoteric descriptive devices, so that the proposed definitions can be perfectly intelligible to the general reader.

(Wierzbicka 1985, 7)

To be sure, the intelligibility of the NSM definitions presupposes that primitives themselves are simple and self-explaining. However, as critics maintain, such definitions have not been sufficiently tested with regard to their comprehension (Geeraerts 2010). As Riemer (2006) remarks, the intelligibility of a dictionary definition depends more on whether defining words are sufficiently familiar to the reader than on whether they represent universal concepts that are insusceptible of conceptual decomposition. Nevertheless, primitives are basic in the sense that they are so frequently used in English that one can hardly imagine communication without them.

Undoubtedly, Wierzbicka’s approach to defining lends theoretical support to the use of a controlled vocabulary in dictionaries. However, there is a huge methodological discrepancy in determining the set of primitives and the set of defining words for dictionaries (Bańko 2006). Primitives are selected on the basis of alleged universal lexicalization, whereas the selection of a defining vocabulary rests, among other factors, upon word frequency and range within a specific language (see Sections 1.5.2 and 2.2.4).

Wierzbicka's conception of primitives has inspired a number of semantic, stylistic and pragmatic studies (Apresjan 1995; Grochowski 1993; Bańko 2001; Żmigrodzki 2003; Croft & Cruse 2004; Hanks 2008). It has also invited much favorable and critical comment in metalexicography (Landau 1993; Hanks 2015). Landau (1993, 113), for example, admires Wierzbicka's method for being "impeccably logical and brilliantly executed", and believes that it has a potential for influencing learner's practical lexicography because it is rich in productively useful information. However, Landau (1993, 116) seriously doubts whether the method can bring benefits for native speakers, as it is "extremely cumbersome, time-consuming, and (wrongly) juvenile". He argues that the method is "far from simple in the demands it places on the reader to follow the course of argument", which is the reason why few readers have the patience to wade through lengthy definitions (116).

There are researchers who doubt that defining exclusively in a specific set of indefinables is practically feasible. Grzegorzczkowska (2001, 57), for example, claims that, when decomposing the meaning of a lexical unit into simple elements, there is always a chance that the definition will require the mention of an element that is unique and specific only to this unit. In such a case, the concept should be treated as indefinable, which inevitably results in the enlargement of the set of primitive concepts. It goes without saying that the search for primitives would be a never-ending process, because new primitives would emerge in new definitions. It may be apt here to anticipate some results of the current research by noting that Grzegorzczkowska's observation sheds light on the fact that learners' dictionaries in fact fail to restrict their DVs to a pre-specified list and use words from outside the list (see Section 4.3.). Her observation also suggests that there may be more than one restricted defining vocabulary, depending on the set of lexemes being defined.

True, Wierzbicka's definitions are lengthy and as such can hardly be applied to practical lexicography, but as Bańko (2001) rightly remarks, they are inspiring for lexicographers and researchers. Other linguists argue in favor of defining words within a set of elementary concepts. Apresjan (1972, 43), for instance, points out that such definitions are more intelligible to the reader. Further, they reveal the hierarchical structure of a language by making explicit the relationship between the meaning of a word being defined and the various elementary meanings activated in a definition (Apresjan, 43). In response to Wierzbicka's suggestion of defining technical concepts without recourse to technical jargon, Landau (1993, 117) admits that defining such words in simple terms would be "revolutionary" and should be given careful thought by commercial lexicographers. Yet he envisages the possibility of using a folk definition along with a traditional scientific one to the benefit of different groups of users.

In turn, Weinreich (1980, 55) maintains that by defining with indefinables one can solve the problem of circularity in dictionaries. He suggests a stratification of a DV into a central core and, like onion rings, into the various levels of the periphery (Allan 1986, 355). The vocabulary of the core would be defined circularly or ostensively (e.g., *blue* “the colour of the sky”), and would be used in definitions of the words of layer 2, which in turn would serve as building blocks for definitions in layer 3. In other words, the vocabulary from more peripheral layers would be defined with the words of the more central layers (Allan, 355). Thanks to this strategy, circularity would be avoided in all the layers but the core one. Weinreich suggests that the core vocabulary might be reduced to the size lower than Basic English, which consists of 850 words. However, he does not specify how many lexical strata should be distinguished or how this vocabulary should be selected (Seretny 1998, 258). The idea of stratification of a DV has been taken up by other researchers, who, for example, considered it as a way of improving existing electronic dictionaries (Rundell 2015, 316), or implemented it in an experimental project (Bullock 2010). In the following section, we will take a closer look at such a project.

2.5.2 A primitive-based dictionary

Bullock devised an experiment to test the “expressive power” of definitions written in Natural Semantic Metalanguage (NSM) (Bullock 2010, 226). To that end, he compiled a small dictionary consisting of definitions of all the keywords in the DV of the *Longman Dictionary of Contemporary English* (LDOCE). He defined the 2,352 keywords using Wierzbicka’s primitives and a few extra words from the LDOCE DV, such as *colour*, *number* and *shape*. Altogether, he used only 64 defining words and phrases (Bullock 2010, 232), imposing as it were a much higher degree of control than is normally found in learners’ dictionaries. The NSM-definitions were meant to be a remedy for circularity. In contrast to lengthy definitions proposed by Wierzbicka, some of which are two-page long, Bullock’s definitions were limited in length to about 100 words. Below is an example of his definition, in which primitives are typed in capitals, and non-primitives appear in normal case.

cup=[noun-only] SMALL THINGS PEOPLE make [to+] contain hot liquid [+s]
 THAT PEOPLE drink [:] ONE PERSON CAN use ONE hand TO_MOVE ONE_OF THESE
 NEAR_TO mouth and CAN drink ALL liquid THAT ONE_OF THESE contain [+s]
 (Bullock 2010, 229)

Bullock tested the quality of his definitions on twelve English-speaking students aged 14–18. They were asked to read a set of NSM-definitions and guess the missing headwords (i.e. the words being defined). The test participants were presented

with a more user-friendly version of definitions than the one exemplified above, with no typographical mark-up (uppercase letters, underscores, the third person present tense morpheme *-s*, brackets, etc.). The participants identified the missing headwords in three tasks. In the first task, they were asked to write each headword in the blank space. The second task was the same but was easier because the initial letter of each headword was already given. In the third task, which was the easiest, the participants were asked to choose each headword from a list. For comparative purposes, the participants carried out these tasks using not only NSM-definitions but also two other definition types: the corresponding definitions in *LDOCE* and the full-length definitions drawn from the NSM literature (Wierzbicka 1996, 1985). The researcher measured the accuracy of each definition type by calculating the percentages of correct responses for each task.

Bullock (2010, 230–1) concludes that although his primitive-based definitions had a lower precision (by approximately 8%) than the other definition types, the meanings they described were conveyed adequately. The experiment was conducted on native speakers, but Bullock (2010) argues that the NSM definitions have a potential of being successfully understood by less competent users of the language, including beginning EFL learners.

Interestingly, however, the effectiveness of primitive-based definitions were questioned by Neubauer (1980, 15–6), who maintained that primitive-based lists are insufficient for lexicographers to write intelligible definitions. Neubauer based his claim on an analysis of various lists of semantic primitives consisting of a very limited number of items (Herbst 1986). As an example, he quotes Ballweg-Schramm and Schumacher's proposal for a definition in a valency dictionary of German, arguing that the definition can hardly be helpful in understanding the word meaning. In their model of the dictionary entry, Ballweg-Schramm and Schumacher offer a definition in the form of a paraphrase (on the right) of the German verb *informieren* (on the left) (Herbst 1986, 115):

Der _{xNom} informiert (den y _{Akk}) (in dem k _{Adv}) (über den z _{Präp})	Dadurch, daß (x) (k) aktualisiert, bewirkt x absichtlich, daß (y) Eigenschaften des (z) erfährt.
--	--

Although the dictionary compilers stress they aim to keep the DV as small as possible, they do not give the details of the list, nor do they reveal the number of items included (Herbst 1986).

Returning to Bullock's project, the researcher suggests that a primitive-based dictionary may serve as a companion to already existing EFL dictionaries, especially for those learners whose proficiency level is lower than expected. He observes that a DV of over 2000 words employed by learners' dictionaries

presupposes that the learner is sufficiently familiar with this vocabulary. In other words, the user is expected to have a receptive vocabulary of that size in order to understand any definition in such a dictionary. Because this may pose a high threshold for less advanced learners, the argument goes, the companion dictionary whose definitions are framed within NSM may help the learner overcome comprehension problems encountered in EFL dictionaries.

Like Weinreich (1980), Bullock suggests a multi-layer structure for the companion dictionary. The first layer would be a list of over sixty primitives. Because primitives cannot be decomposed themselves into smaller units, their meaning would be conveyed by illustrations and translations. The sixty primitives would be used in the second layer to define 700 words selected from the *LDOCE* DV. The third layer would be composed of the remaining 1300 (or more) words from the Longman DV, defined with the vocabulary from the previous two layers (Bullock 2010). The multi-layer structure would ensure a graded progression in learning the English vocabulary, from conceptually simple to complex words. The dictionary would be intended for beginning learners as a means of broadening the receptive vocabulary to the level necessary for successful comprehension of all definitions in *LDOCE*.

The scholar has launched the companion dictionary online with some modifications (Bullock 2019). The first layer is comprised of approximately the same set of primitives as used in the original experiment. As Bullock explains (2019), the primitives represent universal concepts, or “semantic atoms”. There are as many as 34 middle layers, with 300 “semantic molecules”. Each layer has a vocabulary defined with the words of the previous layer. All these 300 molecules plus 61 atoms from the first layer are used in the following layer to define 2000 words of the *LDOCE* DV. Thus, compared to the initial experiment (Bullock 2010), the *LDOCE* DV is here defined using a much larger vocabulary than originally planned (360 vs. 64 defining items respectively). This may have been caused by practical reasons, namely the need to eschew awkward and wordy definitions which inevitably resulted from the reduced size of the DV. As it stands, Bullock’s DV relies on a considerably larger set of words than NSM, but still a significantly smaller set than the DVs in contemporary EFL dictionaries.

Finally, it is worth pointing out that a payoff from defining within a restricted DV is that such definitions are a rich source of information on the language and the world (Wilks et al. 1988). With respect to the degree of semantic detail, they surpass traditional definitions (Piotrowski 2002). The type of knowledge they contain is of special interest to developers of NLP applications. Dolan et al. (1993) used an online version of *LDOCE* to build a lexical knowledge base. The base had a form of a network consisting of tens of thousands of words, which revealed semantic relations between them, such as “hyperonym”, “part of”, “location”, and

“purpose” (Dolan et al. 1993). In turn, Michiels and Noël (1984) used LDOCE in the investigation of the relationship between an instrument (e.g., *instrument*, *tool*, *liquid*) and the process in which it is used. Such projects are of interest to the NLP community.

2.5.3 Minimal English

In 2014 Wierzbicka proposed the so called *Minimal English*. It is a “mini” version of English which “can serve, whenever needed and appropriate, as a common auxiliary inter-language for speakers of different languages, and a global means for clarifying, elucidating, storing, and comparing ideas” (Wierzbicka 2014, 194). Unlike Basic English, Minimal English has been designed to ensure maximum cross-translatability; it is the English version of a conceptual language with its vocabulary representing “the full repertoire of shared human concepts” (Wierzbicka 2014, 195). It is intended to be an auxiliary language that makes it easier to communicate beyond a language barrier. Minimal English owes a great deal to the research into NSM, but it makes use of a larger set of words. The vocabulary of 250 words includes semantic primitives, that is, basic universal meanings which are not themselves definable (see below). In addition, it contains some grammatical words and some words that are “more or less cross-translatable or more or less unavoidable in many contexts of modern, international discourse” (Goddard and Wierzbicka 2018, 8). Examples of the latter category of words are *book*, *buy*, *country*, *government*, *number*, and *sea*. As Goddard argues, “with fewer words to choose from, one is forced to focus on the essential things that one wants to say, without getting distracted by all the available lexical options” (Goddard and Wierzbicka 2018, 6).

2.5.4 Defining with primitives for cross-linguistic comparison

Cross-linguistic comparison represents an interesting area of NSM application, as primitives are universal in the sense of having exact equivalents in all languages. As such they can serve as a means for describing the meanings of lexical units in various languages, a process which requires manual analysis along the lines proposed by Wierzbicka. Essentially, her analysis aims to capture the semantic invariant of a concept in a given language. In order to accomplish this goal, the definer should be able to discover the “internal logic” of the concept (Wierzbicka 1985, 19). Because concepts may differ across languages, this analysis can be used in discovering conceptual differences between two different languages.

Wierzbicka claims that her NSM provides *tertium comparationis* for cross-linguistic comparison (Goddard and Wierzbicka 2008). She believes that explication through reductive analysis can reveal meaning differences between corresponding

lexical items in two languages (Goddard and Wierzbicka 2008). For example, her analysis of the English adjective *green* and its Welsh equivalent *gwyrdd* leads to the conclusion that, although the words are very close in meaning, the latter has a somewhat more restricted range of reference, because it is used to talk about “relatively livelier, brighter, fresher greens” (Wierzbicka 1996, 307). In order to bring out the semantic difference, in the analysis of the Welsh word she suggests a reference to wetness after rain, in order to evoke the associations with “fresh and glistening vegetation”:

X is green. =
 in some places many things grow out of the ground
 when one sees things like X one can think of this
 X is *gwyrdd*. =
 in some places many things grow out of the ground
 at some times there is water in these places (after rain)
 when one sees things like X one can think of this
 (Wierzbicka 1996, 306–7)

Such an explanation may be particularly useful for bilingual lexicographers, because it is capable of showing that even seemingly equivalent words differ in meaning (Adamska-Sałaciak 2010a, 402).

Useful as it is, defining within NSM remains a manual task. It relies on “methodological introspection and thinking” rather than on “laboratory experiments” or “reports of casual, superficial impressions or intuitions (either of ‘informants’ or the analyst himself)” (Wierzbicka 1985, 19). Although the activity of definition writing is time-consuming and is prone to inconsistencies, it is possible to facilitate the process. Wierzbicka offers quite specific guidance on how to perform conceptual analysis. In the case of natural-kind concepts, her definitions systematically cover such aspects as habitat, size, appearance, etc., whereas human artefacts are described in terms of function, material, appearance, and size. Such conceptual frameworks are crucial to a systematic comparison of similar concepts not only within a particular language but also across different languages.

Conceptual frameworks for definitions have been of great interest to researchers of terminology (Sager and Kageura 1994; Sager and L’Homme 1994; Sager and Ndi-Kimbi 1995; Dancette 2005; San Martín and L’Homme 2014; Seppälä 2015). Seppälä (2015), for example, proposes a model for predicting definition contents irrespective of language and subject domain. On the basis of her methodological framework for analyzing definitions, she verifies a hypothesis that the selection of differentiating features is “partially determined by the entity type to which the defined term refers” (45). In an early work by Sager and Ndi-Kimbi (1995), the researchers address the problem of insufficient guidelines for writing terminological definitions. They

explore possible linguistic realizations of conceptual structures of terminological definitions. On the basis of a text corpus, they identify rules and proper patterns for the more systematic creation of definitions. Systematic definitions are designed with a view to modelling knowledge structures and to creating language-independent knowledge retrieval systems.

However, terminological research cannot be applied to general-purpose lexicography without reservations. This is because terminological definitions are limited to the denotation of a lexeme and describe it in terms of necessary and sufficient components of the concept within a particular subject field (Sager and Ndi-Kimbi 1995; de Bessé 1990). By contrast, lexicographic definitions go beyond denotative meaning; they focus on connotation, usage, and syntactic as well as stylistic properties of a lexeme. Such definitions often explain how a word is used and when it is used. General lexicographers are confronted with the problem that meanings of many lexemes have no clear boundaries and thus cannot be described in terms of necessary and sufficient conditions. It is worth recalling here Wittgenstein's search for common elements of referents of *game*, which led him to the conclusion that the referents cannot be described in terms of sufficient and necessary conditions, but rather they can be linked by family resemblance relations (Geeraerts 2006b). Thus, in order to account for different contexts of use of a lexeme, lexicographers search for its prototype, or – in the words of Wierzbicka – the invariant meaning. A consistent comparison of corresponding lexemes in different languages needs a complex conceptual framework that takes into consideration all aspects of meaning and use of the lexemes.

2.6 Summary

After the publication of the GSL (1953), the idea of vocabulary control continued to inspire educators, researchers, and linguists. With advancements in technology and availability of large corpora, it was possible to revise existing controlled vocabularies and compile new ones on the basis of more reliable and representative data. This allowed compilers to apply quantitative criteria more consistently than in the past.

The application of vocabulary control goes beyond teaching and learning. With the development of computer software, vocabulary control received attention from innovators in the field of industry. Vocabulary control was particularly useful for enhancing professional text comprehension and machine translation. Recent practical applications of controlled vocabularies show that the fundamental principles as laid down by the pioneers of the field are still current.

In this chapter we also saw that for some linguists, vocabulary control lies at the heart of lexicographic analysis. Linguistic research offers guidelines for establishing a core vocabulary, which has implications for dictionary makers. To what extent dictionary makers adhere to basic principles of vocabulary control will be the subject of the following chapters.

CHAPTER 3

Defining with a controlled vocabulary: Preliminary insights

The early-20th century interest in vocabulary control led to the compilation of the first EFL dictionaries, with *NMED* (1935) and *ISED* (1942) being major innovations in the field. The dictionaries used different approaches to a DV. The one pioneered by West in *NMED* involved a systematic use of words selected prior to defining, and occasional use of words outside the list. We will see in this chapter that this approach did not strike a chord with Hornby, the compiler of *ISED*, who used no such restrictions in his dictionary.

Neubauer (1989, 900) distinguishes five approaches to controlling a definition vocabulary, from no control to the control of form and meaning:

- a. No control whatsoever over what words are used in the dictionary definitions;
 - b. It is attempted to use 'simple language' in the definitions, but without explicit restrictions;
 - c. The words in the definitions are mainly selected from a specified list of words contained in a defining vocabulary, in addition some words not contained in the defining vocabulary are used for cross-references and terminology; this type of dictionary definition has been called *mixed definitions*;
 - d. All words occurring in the definitions form part of an explicitly compiled defining vocabulary that is listed in the dictionary;
 - e. All words occurring in the explanations form part of an explicitly compiled defining vocabulary with the senses actually used marked, and only the senses listed are used in the definitions.
- Neubauer (1989, 900)

While West's approach corresponds to the letter (c), we can expect that *ISED* and other learners' dictionaries follow at least the approach (b). We can also assume that native-speaker dictionaries control their definition vocabularies to some extent.

The above questions will be dealt with in Sections 3.2 and 3.3. The aim of the sections is to formulate preliminary observations on the issue of controlling definition vocabularies in English dictionaries published up to the early 1980s. I will

continue this line of investigation in more detail in Chapters 4 and 5 by moving on to the more recent period (1984–2015). The final section (3.4) reviews previous studies of the subject, and on the basis of this review, offers several hypotheses that will be verified in the subsequent chapters. Before I begin, I will explain the term *definition*, because it is essential for the study of a defining vocabulary.

3.1 What is the dictionary definition?

In a monolingual dictionary, the definition is the most important component of the entry because its primary function is to give an explanation of the meaning of the headword, or the lexical unit being defined (Atkins and Rundell 2008, 208–209). The need to understand meanings is the main reason for native speakers and learners to consult general-purpose dictionaries (Quirk 1973, Tomaszczyk 1979, Summers 1988). In the words of Wiegand (1994), the dictionary definition is the core part of a semantic comment on the headword, though it is not the only component that serves that purpose. Other components that carry semantic information include illustrative examples, pictorial illustrations, stylistic and attitude labels, as well as synonyms, antonyms, and other elements of paradigmatic information.

Lexicographers use various definition formats, depending on the function of the dictionary, its target audience, and the category of the word being defined. The traditional definition follows Aristotelian logic: it categorises the word being defined (*definiendum*) in terms of a superordinate term (*genus proximum*) and distinguishes it from other words in the category by means of a set of differentiating features (*differentia specifica*) (Svensén 1993, 120–31; Cowie 1999a: 158). This classical approach is suitable for defining nouns, but is less useful in defining verbs and adjectives (Hartmann and James 1998). Below is an example of this type of definition:

headphone a telephone receiver worn in pairs on a headband, esp. for wireless listening (Chambers7)

Another way of defining is by giving one or several near-synonyms (see the entry for **lax** below). In spite of its long-established tradition, defining by synonym is not without controversy among metalexicographers. The so-called synonym definitions are criticized for being vague and misleading for users (Wiegand 1994; Louw 2000) (see also 5.1.2). Firstly, the very fact that absolute synonyms are extremely rare in language undermines their usefulness as a defining strategy. Even when a lexicographer decides to use a near-synonym, it is rarely simpler than the word being defined. Defining a word in terms of words that are frequent and common

across a range of registers and topics requires the use of paraphrase rather than synonymy. Secondly, synonymy is a relation between lexical units (in the narrow sense: between words in their particular single meanings) rather than between ambiguous polysemous words. Synonymy is a complex relation that involves not only denotational but also contextual, stylistic, expressive, and regional differences which exist between lexical units. Recent dictionary practice shows that lexicographers often fail to communicate in which meaning a synonym is used in a definition (Louw 2000). On the other hand, as Landau (2001, 398) argues, synonym definitions “are not necessarily a mark of lazy or inept editing; they may be the best solution to the problem of too little space”. Svensén (1993, 119) adds that the approach is “entirely valid when the need for semantic precision is not too great”.

lax *adj* slack; loose; soft or flabby (Chambers12)

Synonym definitions stand in sharp contrast to full-sentence definitions, as in:

accrue If money or interest *accrues* or if you *accrue* it, it gradually increases in amount over a period of time (COBUILD8)

Introduced in *COBUILD1* in 1987, this definition model is now commonly used in learners’ dictionaries because it is believed to be user-friendly and readily accessible to learners. It resembles the style of definitions that ordinary people use in their everyday conversations. The definition model consists of the introductory clause, which uses the *definiendum* to provide a typical context of use, and the main clause, which is the explanation proper. Thus, the concept of lexicographic definition encompasses a range of definition formats, from a synonym, a series of synonyms, and a brief analytical phrase, to a more discursive account of semantic and contextual properties of the lexical unit.

The varied formats of definitions raise doubts as to their equal status and function in a dictionary entry. Firstly, the full-sentence definition does not furnish purely semantic features of a lexical unit but provides contextual information. Although both of these information types are essential for learners, their usefulness depends on the task at hand (reception or production). Secondly, the synonym definition can be equally treated as an entry component distinct from the definition. This is because the former provides paradigmatic, rather than explanatory, information. Paradigmatic information provided in a dictionary entry does not fulfil the same function as analytical definition because the chief purpose of the former is to satisfy the needs of users for text production, while the latter is essentially used in text reception. Moreover, lexical items, whether presented as synonyms, antonyms, co-hyponyms, or other related items, do not play the same role as definitions with respect to the user’s lexical competence (Seppälä 2016). Their provision in a dictionary entry is an important aid to

vocabulary expansion, whereas analytical definitions are intended to solve comprehension problems by paraphrasing the word being defined. The uncertain status of synonym definition is exacerbated by the inadequate cross-referencing of synonyms in entries, for example between the entry where a synonym appears along with an analytical definition and the entry for this synonym (Louw 2000). Although in many dictionaries, defining a word by a brief analytical phrase along with a near-synonym is a long-standing practice (see the entry for **debase** below), metalexigraphers (e.g., Wiegand 1994) believe that such a presentation of synonym information is misleading for users because it fails to explain the word meaning.

debase to make mean or of less value: to adulterate

(Chambers7)

The disjunction between lexicographic practice and theory leads to at least two different interpretations of the term *definition*. According to Burkhanov (1998), the term represents at least two concepts. The first has a broad meaning and refers to a combination of a word being defined (*definiendum*) and the explanation of the meaning of this word (*definiens*). Such an interpretation encompasses the full-sentence approach to the explanation of meaning, as used in *COBUILD*. This means that if-clauses in *COBUILD* definitions qualify as parts of definitions. A second concept designated by the term *definition* applies only to the defining part that explains semantic and lexical properties of the *definiendum*, while excluding its syntactic properties (Burkhanov 1998). Thus, applying the concept to the *COBUILD*-style definitions, one concludes that only the main clause meets the criterion for definition.

Both interpretations are applicable to the analyses in the present book. In a large part of the book, I will use the term in the former broader sense, but restrict its use to the narrower sense for the analysis of receptive vocabulary load of definitions in Chapter 4. This restriction is motivated by the rigor of quantitative analysis of the vocabulary (see 4.1.2). I will follow typographical conventions of dictionaries under study and apply the term *definition* to what is traditionally known as synonym definitions, while excluding all the synonyms that are explicitly marked as separate components of a dictionary entry.

The definition formats presented in this section (analytical definition, synonym definition, and full-sentence definition) are used in general-purpose lexicography for their practical values. If we assume that the ideal dictionary definition is the one that is easily accessible to the target user, then it may be close to the full-sentence definition. However, accessibility is not the only quality that dictionary definitions should possess. As Johnson argued, the ideal of definition combines “brevity, fulness, and perspicuity” (Johnson 1747). Such a compromise may be more difficult to achieve than it seems. The following section looks at the history

of English lexicography to see how much attention English lexicographers paid to accessibility of their definitions.

3.2 Defining in plain English: A historical overview of native-speaker dictionary tradition

As mentioned earlier, Johnson (1747) believed that one of the key features of the ideal definition is its perspicuity. The requirement of clarity applies to learners' dictionaries as well as native-speaker dictionaries. The history of the latter dictionaries deserves our attention not only because of the topic of definition vocabulary but also because of the fact that they bequeathed many of their features to the early learners' dictionaries.

The notion of simplicity can be traced back to the beginnings of the English monolingual lexicography. The early monolingual lexicographers of the 17th century explained difficult words using, what they referred to as *plain English*, that is, the language which was supposed to be "straightforward and easy to understand"¹ for the target user (McArthur 1998a, 109). By using plain English in definitions, the lexicographers conformed to the fundamental principle of dictionary making, namely that difficult words should be explained with easy ones. This section seeks to review how the idea of plain English was implemented in major English dictionaries. After this review, I will focus on how this idea gained attention from readers of publicly available texts.

The early monolingual dictionaries of the 17th century, such as Cawdrey's *Table Alphabeticall* (1604), Bullokar's *An English Expositor* (1616), Cockeram's *The English Dictionarie* (1623), and Phillips' *The New World of English Words* (1658), belong to the "hard word" tradition of English lexicography. These dictionaries had as their aim to explain the poorly anglicized words of classical origin with the familiar words of vernacular English (Stein 2002a). Hard words, which became a hallmark of learned writing, had flooded the English language since the Renaissance period. The idea that these difficult words could be explained in plain English appealed to the middle classes who lacked classical education (McArthur 1998a, 134). The notion of plain language was clearly synonymous with that of familiar vernacular.

1. The OED (3rd ed.) defines **plain English** as "plain straightforward language", and the adjective **plain** as "free from ambiguity, evasion, or subterfuge; straightforward, direct". The OED's quotation (from 1693) containing *plain English* postdates the use of this combination in Cawdrey's *Table Alphabeticall* (1604).

Cawdrey's *Table Alphabeticall* is an early attempt at explaining difficult classical terms with familiar English counterparts. His major technique of defining was using one or more near-synonyms, with occasional brief analytical phrases (see below). Given the predominant use of synonym definitions, *The Table* resembled the bilingual dictionaries of the 16th century, in which foreign words had been paired with native ones.

alien, a stranger

colleague, companion

colume, one side of a page of a booke

violent, forcible, cruell, iniurious

(Cawdrey 1604)

The compilers of the early monolingual English dictionaries made various assumptions about the usefulness of their works for various groups of users and for different situations of dictionary use. Cawdrey, for example, explained the pedagogical function of *The Table Alphabeticall*, emphasizing that the dictionary was designed to teach words both productively and receptively:

conteyning and teaching the true writing, and understanding of hard vsuall English wordes, ... With the interpretation thereof by plaine English words, gathered for helpe of Ladies, Gentlewoman, or any other vnskilfull persons. Whereby they may the more easilie and better understand many hard English wordes, which they shall heare or read in Scriptures, Sermons, or else where, and also be made able to vse the same aptly themselues ...

(Cawdrey 1604)

In a similar fashion, Phillips believed that his *New World of English Words* (1658) was “very necessary for strangers, as well as our own countrymen, or for all persons that would rightly understand what they discourse, write, or read”. Later in 1708, Kersey published *Dictionarium Anglo-Britannicum*, declaring that it was designed “for the benefit of young students, tradesmen, artificers, foreigners, and others who are desirous thoroughly to understand what they speak, read, or write”. Interestingly, the title page of Bailey's *Universal Etymological English Dictionary* (1721) bears a declaration which closely resembles that given by Kersey in his *Dictionarium*, but with the word order slightly changed: “for the benefit of young students, artificers, tradesmen and foreigners, who are desirous thorowly to understand what they speak, read, or write.”²

2. This fragment can be considered as an instance of copying dictionary content, or collating material from predecessors (McConchie 2013, 118), which was commonplace among lexicographers of that period (Podhajecka 2009, 157). As can be seen, copying was not confined only to the content of entries but extended to title pages.

While there is no doubt that these early dictionaries developed in response to practical needs of their users, the lexicographers' assumptions were certainly not informed by an in-depth analysis of requirements of the real user. These early dictionary makers, however, were not unlike a number of later lexicographers who acknowledged various functions of their dictionaries, but rarely had sufficient evidence to support their claims about the target user's needs (Hartmann 1989, 123). While the hard-word dictionary evolved into the more comprehensive reference book of the 18th century, and then into the multi-volume encyclopedic dictionary of the 19th century, lexicographers attached little interest in such matters as how dictionaries were really used by their users and what to do to make definitions more accessible to immigrants and other learners who were non-native speakers of English (McArthur 1998a). This state of affairs persisted into the 20th century.

In the 18th century, hard-word dictionaries became more inclusive, slowly evolving into a record of the entire lexicon. In his *Universal Etymological English Dictionary* (1721), Bailey included not only hard words, many of which he took from the dictionaries already published, but also dialect words, cant, and obsolete terms used in the English literature (Landau 2001, 63). Dictionaries began to record everyday words and meanings, which raised a problem of how to define them. Defining a common word in terms of more common ones was not always feasible; more importantly, however, it does not seem to have been practical, given the fact that they were known to native users. In Bailey's *Universal*, definitions of common animals were hardly more than a brief indication of the class to which the animal belonged, giving only a gist of the word meaning; for example:

Horse ... a Beast well known.

Mouse ... a Domestic Creature well known. (Bailey 1721)

Considering the superficiality of these definitions, one may wonder why they were given at all. Interestingly, foreigners belonged to the target readership, as the title page explicitly indicates:

The whole work compiled and methodically digested, as well for the entertainment of the curious, as the information of the ignorant, and for the benefit of young students, artificers, tradesmen and foreigners, who are desirous thorowly to understand what they speak, read, or write. (Bailey 1721)

Admittedly, the above definitions were unhelpful for learners in both receptive and productive mode. Therefore, it is probable that they were given just for the sake of inclusion, as comprehensiveness and dictionary size had become important selling points in dictionary business (Béjoint 2000, 95).

In 1755 Samuel Johnson published *Dictionary of the English Language*, the first English dictionary with definitions systematically supported with citations

from literary “authorities” and “the best authors” (Johnson 1755). Initially, the dictionary was envisaged as a highly prescriptive work, in which the literary evidence would serve as a standard of the language.³ According to Hanks (2005, 243), Johnson’s definitions were “terse, stylish, and sometimes witty, as well as factual, clear, and comprehensible (with a few exceptions, some of them deliberate).” Johnson’s dictionary has been admired not only for the quality of definitions and the consistent use of quotations, but also for his thorough understanding of lexicographic problems he was confronted with while working on the dictionary. The compiler recognized the need to write definitions using “terms less abstruse than that which is to be explained” (Preface, Johnson 1755). However, he was well aware of difficulties in finding such terms:

... such terms cannot always be found; for as nothing can be proved but by supposing something intuitively known, and evident without proof, so nothing can be defined but by the use of words too plain to admit a definition.

(Preface, Johnson 1755)

The following passage encapsulates problems for lexicographers compiling monolingual dictionaries:

To interpret a language by itself is very difficult; many words cannot be explained by synonymes, because the idea signified by them has not more than one appellation; nor by paraphrase, because simple ideas cannot be described. When the nature of things is unknown, or the notion unsettled and indefinite, and various in various minds, the words by which such notions are conveyed, or such things denoted, will be ambiguous and perplexed.

(Preface, Johnson 1755)

Johnson argued that the words that signified what he called after Locke “simple ideas” are impossible to define. The words would now be considered as semantically simple. Using Wierzbicka’s terminology, we would perhaps say that the words are non-decomposable, that is, their meaning cannot be divided into smaller semantic parts (Wierzbicka 1996, 237). Johnson was skeptical about defining “the easiest words” and believed that such definitions lead to circularity:

... many seeming faults are to be imputed rather to the nature of the undertaking, than the negligence of the performer. Thus some explanations are unavoidably reciprocal or circular, as /hind, the female of the stag/; /stag, the male of the hind/: sometimes easier words are changed into harder, as /burial/ into /sepulture/ or /

3. In the initial stage, Johnson intended to collect excerpts from literary “authorities” as evidence for his pre-established categories of meaning. However, in the course of the work, he found out that the actual usage did not fit neatly into his system of categories (Reddick 2009). This discovery led him to the conclusion that language cannot be prevented from change.

interment/, /drier/ into /desiccative/, /dryness/ into /siccity/ or /aridity/, /fit/ into /paroxysm/; for the easiest word, whatever it be, can never be translated into one more easy. (Preface, Johnson 1755)

A much quoted definition of **network** illustrates that even common words may need to be defined in terms of rare and abstruse words:

network Any thing reticulated or decussated, at equal distances, with interstices between the intersections (Johnson 1755)

Johnson assumed that his dictionary would also be useful for non-native speakers of English. When defining by synonym, he deliberately used a pair of words of different origins: Germanic and Romance. He used both words in the hope that one of them would be easily interpreted by the foreign learner by virtue of being similar in form to the learner's native cognate word. Such double definitions were introduced with the learner's needs in mind even though the Romance words might have been obscure to the native user:

But easiness and difficulty are merely relative, and if the present prevalence of our language should invite foreigners to this dictionary, many will be assisted by those words which now seem only to increase or produce obscurity. For this reason I have endeavoured frequently to join a /Teutonic/ and /Roman/ interpretation, as to cheer to /gladden/, or /exhilarate/, that every learner of /English/ may be assisted by his own tongue. (Preface, Johnson 1755)

The ideal of the English dictionary as popularly perceived in the 19th century was a multi-volume encyclopedic work (Simpson 1990). In the United States, this ideal was embodied by Webster's *American Dictionary* and its revisions (1828 and 1864). Webster aimed at comprehensiveness of coverage and thoroughness of treatment. He heavily drew on Johnson, but aimed to surpass the predecessor by covering a number of terms of developing art, science, and technology, and using a fine-grained division of senses. According to Friend (1967, 36), his definitions were "more accurate, more comprehensive, and not less carefully divided and ordered than any previously done in English lexicography". Many of his definitions were produced by copying Johnson's definitions verbatim and incorporating the information from Johnson's citations into the definitions.⁴ Webster's dictionary abounded with biological and geographical

4. For example, in Webster's definition of **nice** "Requiring scrupulous exactness; as a nice point" (1828), the lexicographer has copied Johnson's definition "Requiring scrupulous exactness" (1785) and added "a nice point" (Kamiński 2019). This addition comes from Johnson's citation under the same entry: "Supposing an injury done, it is a nice point to proportion the reparation ... *L'Estrange*."

information and other encyclopedic features, many of which were added during the war with the competing Worcester's *Dictionary of the English Language* (1860) in the belief that they would enhance the commercial value of the product (Morton 1994, 60). Scientific words were defined with encyclopedic precision, which resulted in lengthy definitions that were interspersed with scientific, technical, and rare terms.

The prevailing scientism and historicism of the 19th century underlay the creation of other large-scale dictionaries. On the American side, Webster's *American Dictionary* (1864) was followed by other dictionaries with rich encyclopedic content, including Webster's *International* (1890), Whitney's six-volume *Century Dictionary* (1889–1891), and Funk and Wagnalls *Standard Dictionary* (1893–1895). In Britain, the encyclopedic tradition was followed by Ogilvie's *Imperial Dictionary* (1850) and Annandale's revised edition of this dictionary (1882). The concern for deficiencies of English dictionaries inspired British lexicographers to launch a project for a scholarly dictionary based on historical principles. The *Oxford English Dictionary* (1882–1928), published originally in ten volumes, is regarded as the greatest lexicographic enterprise of the period. All the above large-scale dictionaries, completed or still in the making, provided a source material for smaller portable dictionaries that would serve the needs of the growing masses of people.

The nineteenth-century social and economic changes created the need for a popular dictionary. The growth of the middle classes in power and prosperity coincided with their increased state of literacy. The expansion of the education system in Victorian Britain, the establishment of public libraries,⁵ and the intense circulation of papers⁶ had far-reaching consequences for the members of the rising middle class, who had much easier access to printed matter than ever before. The dictionary had an important role to play in the educational development of the people: it served as an aid in reading a wide range of texts available to them.

5. The law regulating establishment of libraries in England was enacted in 1850.

6. Circulation of newspapers in Britain increased rapidly in the course of the 19th century. The reasons can be traced back to the 18th century, when a series of legislative decisions resulted in lifting the barriers to publishers. In 1774 the copyright law was enacted, which made it easier for publishers to issue printed texts without fear of being sent to prison (Blake 1996, 273–4). As a result, publishers had more freedom in expressing their opinions (Blake, 273–4). A few decades later, in 1835, the stamp tax on newspapers was reduced significantly only to be abolished in 1855. The changes had far-reaching consequences, as more and more newspapers were circulated at a much reduced price. These new circumstances contributed to the widespread access of the public to reading matter (Blake, 274; Reisner 1925, 245).

While the dictionary was becoming “a very marketable commodity amongst the self-taught” (Simpson 1990, 1960), publishers set out to reach the emerging mass market by producing dictionaries for the new audience. The idea was to design dictionaries that would be both affordable and accessible to the target user. There was a good deal of experimentation with new formats of presentation of lexicographic material in the hope of finding the one that would appeal to people at large (cf. Simpson 1990; McArthur 1986, 134). The 1872 edition of *Chambers* used a simplified notation of pronunciation, unlike the one used in the parent dictionaries, *Webster-Mahn* (1864) and *Worcester* (1860) (Kamiński 2013). *Chambers* (1872, v) aimed to provide “within reasonable limits, and at a moderate price” the vocabulary that was sufficiently wide to satisfy the needs of the general user. While this early edition presented the vocabulary in strict alphabetical order, the 1898 edition aimed at greater comprehensiveness of coverage by extensive nesting of derivatives, compounds, and phrases, which were printed in large bold typeface. This format of data presentation, which had arguably been taken over from Annandale’s *Concise Dictionary of the English Language* (1886), would prove extremely popular in the first half of the 20th century (see below). Collins published dictionaries in “Clear-type” to achieve visual simplicity (Simpson 1990, 1964). In one of them, *Graphic English Dictionary* (1903, Preface), the headwords were distinguished by being printed in large bold type so as to “lessen the trouble of consulting” the dictionary. Visual simplicity, or the clarity of information presentation, was a recurring theme in dictionaries for the mass market. However, it was hardly perceptible in actual definitions. Clarity of definitions was difficult to achieve in dictionaries in which space was at a premium.

Innovations were also fostered in large-scale dictionaries. For example, the unabridged *Standard Dictionary* (1893–5) broke a few lexicographic conventions in order to achieve greater accessibility of information. The editor placed “the most common” meaning first, because – as he believed – it was the one that the “average user” commonly sought. In addition, etymology appeared at the end of the entry on the grounds that “many persons are confused and discouraged when the signs and abbreviations, and the foreign or other strange words, that make up the etymology, stand between the vocabulary word and its definitions.” (Funk 1893–1895, xi, xii; cf. McArthur 1986, 134). The editor inserted synonyms, wherever possible, after the analytical phrase, because, as he remarked, “[f]ew words, if any, are exact synonyms of others; hence to define by synonyms is confusing” (Funk 1893–95, xi). These decisions were informed by the editor’s assumptions about how the popular audience used reference works.

However, it was the principle of conciseness that radically changed the shape of popular dictionaries. This principle was adopted to the extreme in *CODI*

(1911), compiled by the Fowler brothers. As the editors explained, space was saved by using “the severest economy of expression – amounting to the adoption of telegraphese – that readers can be expected to put up with” (*COD1*, vi). A characteristic feature of *telegraphese* was the frequent use of ellipsis, abbreviations, and commas, to compensate for the severe conciseness of the metalanguage (Ilson 1986).

COD was an immensely successful and influential dictionary. It was drawn on by Hornby in the compilation of his two learners’ dictionaries: the *English Reader’s Dictionary* (1952) and the *Idiomatic and Syntactic English Dictionary* (1942). Hornby explicitly gave credit to *COD* in the prefaces to these dictionaries. In *ISED*, the editor acknowledged dependence on *COD* as far as hyphenation in compounds is concerned, but a brief comparison of definitions from the two dictionaries confirms *COD* as “the source of much in *ISED* that lies outside the core vocabulary of IRET research” (Cowie 1999a, 47). *ISED* definitions inherited from the source dictionary a number of difficult rare words and ellipsis, though Hornby questioned the suitability of the Fowlers’ *telegraphese* for foreign learners (Cowie 1999a).

Rare terms used in *COD* definitions are easily traceable to the parent dictionary, *OED*. A comparison of the definitions between these dictionaries reveals that the two works share the same or similar words:

burn in impress indelibly	(<i>COD1</i>)
burn in to render indelible (the painting upon pottery, etc.)	(<i>OED</i>)
bummer Idler, loafer	(<i>COD1</i>)
bummer An idler, lounge, loafer	(<i>OED</i>)
betray Give up treacherously (person or thing to enemy)	(<i>COD1</i>)
betray To give up to, or place in the power of an enemy, by treachery or disloyalty	(<i>OED</i>)

An inspection of other dictionary pairs reveals similar findings. For example, the editor of *Chambers’s Etymological Dictionary* (*CED*), a dictionary intended for school students and a general audience, declares that “[t]he greatest care has been taken to express the meanings in the simplest language” (1867, v). However, for the lexicographer, the concept of simplicity was clearly synonymous to brevity and conciseness, and he made little attempt to make definitions lexically simple. A closer comparison of random definitions between *CED* and the source dictionary, *Webster-Mahn* (1864), shows that the former avoided the syntactic complexity of the source definitions but retained the same obscure words. For example, the definition of **parson** below is a truncation of the corresponding definition from *Webster-Mahn*. It is clear that the *Chambers* editor has borrowed the words that

carry the most important semantic load (mainly category words) and reused them in his definition:

parson the priest or incumbent of a parish; a clergyman (Chambers 1867)

parson 1. The priest of a parish or ecclesiastical society; the rector or incumbent of a parish, who has the parochial charge or cure of souls. 2. A clergyman; one who is in orders, or who has been licensed to preach.

(Webster-Mahn 1864)

Popular dictionaries employed traditional defining strategies using brief analytical phrases and strings of near-synonyms. In order to save space, they used a range of defining formulae to “account for contextual variability within a single, concise defining statement” (Atkins and Rundell 2008, 438). Here are some examples from the early popular dictionaries:

dyspeptic Afflicted with, pertaining to, or arising from indigestion
(Chambers 1872)

clausular pertaining to, or consisting of, a clause or clauses (Chambers 1901)

recoil act or fact or sensation of recoiling (COD1 1911)

triphthongal Of, pertaining to, or being, a triphthong (Collegiate 1910).

Atkins and Rundell (2008, 438) quote similar examples from more recent dictionaries, pointing out that such formulae are surprisingly long-lasting in English lexicography:

wolf any of various large dog-like mammals ... (Third New International 1961)

abstinence the action or practice of abstaining (OED2 1989)

absurdity the state or quality of being absurd (OED2 1989)

pedantic of, relating to, or being a pedant (Collegiate11 2005)

As the researchers explain, the definition of *pedantic* above is brief yet informative. For the user who already knows the meaning of *pedant*, which is not a common word, it provides many linguistic facts. The participle *being* accounts for the use of *pedantic* in the context of a person, as in the phrase “don’t be pedantic”, and the prepositions *of* and *relating to* in the context of abstract nouns, as in “pedantic neatness”. Other formulae typical of native-speaker dictionaries include “any of various”, “a state of being”, “quality of being”, “the act of *-ing*”, “of or pertaining to a –”. As Atkins and Rundell observe (2008, 438), such formulae are undoubtedly useful for lexicographers as they cover a range of contexts within limited space, but their usefulness to the user is arguable. Nuances of word usage and semantic preferences are left implicit. Incidentally, the word *quality* which pervades dictionary definitions is not even used in its basic or core meaning (Atkins and Rundell 2008, 448–9). It occurs in the string “quality of being”, which is strictly speaking a lexicography-specific construction recorded only twelve times in the British

National Corpus (Atkins and Rundell 2008, 448–9). In this corpus, *quality of* is usually used in the basic meaning, that is, when talking about how good or bad something is, as in: *improving the quality of medical care* (Atkins and Rundell 2008, 448). The above formulae contribute to the brevity of definitions, even though it is achieved at the expense of comprehensibility.

An unrestricted defining vocabulary is another long-lasting feature of popular dictionaries. Below are a few examples of definitions from the early and mid-1980s. In all of them, technical accuracy and precision outweigh the need for simplicity of the vocabulary.

- chalk** white soft rock, composed of calcium carbonate, chiefly shells of Foraminifera (Chambers 1983)
- feather** any of the light outgrowths that form the external covering of the body of a bird and consist of a shaft bearing a set of barbs on each side that interlock to form a continuous vane (LDEL 1984)
- feather** one of the light horny epidermal outgrowths that form the external covering of the body of birds and that consist of a shaft bearing on each side a series of barbs which bear barbules which in turn bear barbicels commonly ending in hooked hamuli and interlocking with the barbules of an adjacent barb to link the barbs into a continuous vane (Collegiate9 1983)
- fencing** The art of using a foil, épée, or saber ... (AHD 1982)
- horn** Non-deciduous outgrowth (COD7 1982)
- kinkajou** a slender nocturnal arboreal carnivorous mammal (Potos flavus, family Procyonidae) of Mexico and Central and So. America that is about three feet long and has a long prehensile tail, large lustrous eyes ... (Collegiate9 1983)
- pheasant** any of numerous large often long-tailed and brightly colored Old World gallinaceous birds (*Phasianus* and related genera of the family Phasianidae) many of which are raised as ornamental or game birds (Collegiate9 1983)
- thyroid** ... a ductless gland in the neck whose overactivity may lead to exophthalmic goitre, and defect to cretinism (Chambers 1983)

These definitions indicate that the basic lexicographic principle of defining in plain and unambiguous language was difficult to follow consistently in popular dictionaries throughout much of the 20th century. Arguably, simplicity of definitions was taken for granted.

Plainness in language is a quality desired not only by lexicographers and linguists (see also Section 2.5) but also ordinary people. Perhaps the best example of people's interest in linguistic simplicity is a movement to promote plain English in official texts (Aitchison 1986; Bennett 1993; Eunson 1996; McArthur 1998b; Williams 2004). The texts in question include documents, reports, notifications, laws,

regulations, and other texts used in public communication. The pragmatic needs for textual clarity and comprehensibility prompted a group of people under the leadership of Chrissie Maher to establish the Plain English Campaign. Set up in 1979, the organization fights against “gobbledygook, jargon and misleading public information” (*Plain English Campaign* 2019). The Campaign hands out annual awards to official bodies in several categories, including Plain English and Golden Bull, the latter for the most confusing verbal gobbledygook. For example, in 1997 the Golden Bull award went to the Department of Trade and Industry for a definition of **pram**, which had appeared in *The Wheeled Child Conveyances (Safety) Regulations 1997*:

‘perambulator’ means (subject to paragraph (2) below) a wheeled vehicle designed for the transport in a seated or recumbent position of one or two babies or infants who are placed inside a body of boat- or box-like shape, but does not include any carry-cot or transporter for any carry cot.

(cited after Page 2001, 124)

A more extreme case revealed by Maher was a 229-word definition of *bed*, which was published by the National Health Service of Great Britain (Redmond 2010). When commenting on the instances of bureaucratic jargon, Maher pointed out that “[b]y poking gentle fun at these examples, we hope to encourage writers to think more carefully about the language they use.” (*Plain speaking is no joke* 1997). As McArthur (1998b, 116) remarks, the Plain English Campaign is a grassroots initiative of “extraordinary ‘ordinary’ people” who, by attracting the attention of the media, put pressure on the government and institutions to avoid unintelligible jargon, technical terms, and unnecessarily complex sentences when addressing general audiences.

The Plain English Campaign reminds one that public institutions are responsible for maintaining linguistic standards when providing services to people (Crystal 2006). This trend must have had some influence on the publishers of popular dictionaries; after all contemporary dictionary compilers are now more sensitive to the needs of the target readership than they used to be at the turn of the 20th century.

However, as shown in this section, plainness in definitions has not always been easy to achieve in dictionaries for a wide readership. Some of the dictionaries published in the early 1980s, especially those belonging to the American tradition, show a marked tendency to define words with scientific precision. This is not only seen in definitions of purely scientific and technical terms, but also the names of species (e.g., *pheasant*), as well as the common names for abstract and concrete concepts (e.g., *chalk*, *feather*, *horn*, *network*). Perhaps publishing houses assume that scientific accuracy is what their target users need. It seems, however, that

recently the demand for encyclopedic information in dictionary definitions has decreased in line with the accessibility of encyclopedic material on the Internet, so one may expect that the latest dictionaries limit the amount of this information.

While the inclusion of encyclopedic information in definitions presupposes that a native user is familiar with specialist terms, lexicographers compiling learners' dictionaries cannot make such presumptions. It would be unreasonable to expect that a learner will know rare and technical, albeit precise, terms used in definitions. Simplicity of a defining language is the fundamental principle of EFL lexicography, and lay at the heart of the first learners' dictionaries: *NMED* and *ISED*.

3.3 Defining in learners' dictionaries between 1935 and 1986

The need to define in plain English is a natural consequence of the trend in foreign language pedagogy to teach through the medium of a foreign language. Known as the direct method, the trend originated in the late 1880s mainly in France and Germany (Cummins 2008, 66). It arose in opposition to the grammar-translation method, which had been the standard school of teaching of classical and modern languages for centuries (Howatt 2006, 642; Howatt 1984; Cummins 2008, 66). The direct method aimed to avoid the shortcomings of the grammar-translation method, such as the necessity of memorizing grammatical rules, use of unnatural sentences for translation, and focus on written rather than spoken language (Howatt 2006). The grammar-translation method legitimized the widespread use of bilingual dictionaries in foreign language education (James 2003, 135). Michael West and James Endicott looked critically at this type of dictionary:

Every time a child refers to an English Vernacular dictionary his mind is switched out of English and he is encouraged to translate instead of thinking aloud" (*NMED* 1941, 1, cited in James 2003, 135)

According to James (2003, 136), the major driving force behind West's involvement in lexicography was his dissatisfaction with bilingualized English-Bengali dictionaries, which his students were using in Bengal in the 1920s. What West perceived as a serious flaw of this type of dictionaries was that they promoted students' use of their mother tongue. As a lexicographer, he wanted to eliminate this shortcoming by removing native language equivalents, and to encourage the use of English by simplifying the language of definitions (James 2003). The design of his new dictionary was consistent with the direct method of language teaching, which used a foreign language as the medium of instruction. Just as the direct method advocated that grammatical rules and any other intricacies of a foreign language

be explained in this language (Bhatia 2006, 21), so the EFL dictionary explained meanings in English. However, it was obvious that for a learner to benefit from such a dictionary, the definitions had to be comprehensible. A restricted defining vocabulary was the means to achieve that end. As Cowie (1999b, 3) points out, it would “be no exaggeration to say that vocabulary limitation gave birth to the learner’s dictionary” (see also Rundell 2008a, 223).

As shown in Section 3.1., the idea of vocabulary control was completely novel to lexicography in the early decades of the past century. Until then dictionaries had been compiled with native speakers in mind, and little consideration had been given to the possibility that a dictionary would be used by second or foreign language learners (cf. McArthur 1998a, 136). Although some dictionary compilers were aware of the fact that their dictionaries would be used by less competent users of the language, they practically ignored receptive needs of this group of users. The *NMED* (1935) was the first dictionary in which the needs received due attention (McArthur 1998a, 136).

3.3.1 Early learners’ dictionaries: *NMED* and *ISED*

In the words of Rundell (2008a, 223), *NMED* and *ISED* embodied the “fundamentals of a monolingual learner’s dictionary.” Both dictionaries in a broad sense controlled the vocabularies of their definitions, in order to make them easily accessible to the learner. However, they used entirely different methods to achieve that effect.

3.3.1.1 *NMED* (1935)

Compiled by West and Endicott, *NMED* (1935) was the first dictionary specifically designed for foreign language learners, and the first to adopt a restricted DV. This vocabulary was not appended to the dictionary, but was described in detail in West’s *Definition Vocabulary*, which appeared in the same year (see Section 1.3.2.1). The dictionary was intended for the learner at the intermediate level as an aid in reading rather than writing. *NMED* provided no syntactic patterns, while inflectional information was given in a rather unsystematic way (Cowie 1999a). It was designed as a companion to West’s simplified readers (see further below). In order to economize space, restrictions were also imposed on the selection of the vocabulary covered “by omitting the rare and highly technical words which the foreigner is unlikely to meet” (*NMED*, iii).

Assuming that the meaning of some words is inferable from their parts, West included in the DV the commonest affixes (e.g., *dis-*, *in-*, *-able*, *-en*). Thus, the learner could guess the meaning of verbs denoting actions such as *harden*, *soften*, *roughen*, as they were regularly derived from the verb by means of the suffix *-en* (West 1935, 16). To some extent, the success of this strategy depended on the

learner's ability to infer the meaning, and the task was easy as long as the word was semantically transparent. West did not ignore this fact, but he was not always successful in achieving full transparency (Cowie 1999a, 25). As Cowie (1999a, 25) observes, *happening* is more than "something that happens", and *harbour-master* cannot be interpreted simply as "the master of a harbour".

Occasionally, West permitted words outside the defining vocabulary. They appeared in italic type often followed by a parenthetical gloss. This resulted in what he called double definition, for example:

trump ... the sound of a *trumpet* (=brass instrument of music)

gherkin ... a small, long and round green vegetable used for making *pickles*
(=vegetables boiled in *vinegar* = a very sour liquid)

Some definitions consist exclusively of partial synonyms, a trace of native-speaker dictionaries:

hard ... difficult; not soft; unkind, cruel or unjust

Reading definitions, one cannot help noticing some untypical uses of words with peripheral, metaphorical or idiomatic meanings. For instance, the preposition *under (him)* (in the definition of **hardmaster** below) denotes a relation of subordination in the workplace, rather than, what is typical, a spatial relation:

A hard master = one who forces those under him to work too much

Another difficulty is posed by a nominal sense of *wrong*:

accuse ... to say that a person has done wrong, to blame

Other sources of difficulty for the learner may be a nominal sense of *gain* and the idiomatic phrase *to be in high spirits*:

A hard bargain = an agreement which will bring little gain;

To be in high feather = to be in high spirits

The problem of the words used in peripheral meanings or opaque combinations will persist for decades in learner's lexicography. They present the areas for improvement in subsequent dictionaries.

3.3.1.2 *ISED (1942) and ALD1 (1948)*

Compiled by A.S. Hornby, E.V. Gatenby, and H. Wakefield, *The Idiomatic and Syntactic English Dictionary (ISED)* was intended specifically for Japanese students of English at the advanced level.⁷ The book was brought out in Japan by the publisher

7. On the basis of ALD1, Hornby compiled two other dictionaries for learners, both in collaboration with Parnwell. They were published in the same year, 1952, and both were intended

Kaitakusha in 1942. It was a seminal work, which earned Hornby a prominent position in EFL lexicography (Landau 2001, 74). The dictionary was reprinted in 1948 without any alterations by Oxford Publishing Press, under a different title: *A Learner's Dictionary of Current English* (ALD1). In 1952, the title was changed again into *The Advanced Learner's Dictionary of Current English*.

Like NMED, *ISED* provided definitions in English, in accordance with the pedagogical principle of teaching a foreign language through the medium of this language. Hornby's research on syntactic patterns and his career as a teacher show that he was committed to this principle. As far as teaching word meanings is concerned, he recognized a few procedures advocated by IRET, such as ostensive and contextual methods and definitions in simple language (Cowie 1998b, 258). According to Hornby, translation should generally be avoided, unless the other procedures are "difficult, impossible or wasteful of time" (Hornby 1938, 22). As Cowie (1998b, 258) remarks, Hornby treated translation as a "procedure of last resort". Given Hornby's commitment to the direct method, it may come as a surprise that he published, jointly with his Japanese colleague Ishikawa, *A Beginners' English-Japanese Dictionary* in 1949. It is not a widely-known dictionary (Cowie 1989b, 258). Hornby was against using learners' mother tongue as the initial means of teaching new words (Cowie, 258), and his dictionary was intended merely as "a reminder of what has already been learnt, but for the moment forgotten" (Hornby 1938, 22).

ISED is praised as a dictionary that broke new ground in production-oriented lexicography (Cowie 1999a, 65). It represents the fruition of the work on syntactic patterns and illustrative examples, undertaken in the preceding years by Hornby, West, and Palmer (Rundell 2008a). It catered to the encoding needs of learners by providing a great number of word combinations, including idioms and collocations. It offered a verb pattern system, which was more elaborate and fine-grained than Palmer's, on which it was modeled (Rundell 2008a, 223).

In the area of vocabulary restriction, Hornby adopted quite a different approach from West's. Hornby intended to make *ISED* definitions "as simple as possible", but he used "no special definition vocabulary" (ALD1, v). As he explained, "the compilers could have no confidence that the definition vocabulary would be known to the prospective users of the dictionary" (ALD1, v).

for the learner at a lower level of proficiency. According to the editors, *The Progressive English Dictionary* was designed to help "a student who has completed an elementary course in English language to read books written in everyday English and to widen his vocabulary", whereas *An English Reader's Dictionary* was intended to "render the same service to intermediate students engaged in wider reading and thus needing a larger vocabulary" (Hornby and Parnwell 1952b, v). While the former was designed primarily to meet the decoding needs of the user, the latter catered for both decoding and encoding needs, as it offered illustrative examples.

Nevertheless, the compiler aimed to adhere as much as possible to the principle of defining by common words:

[Defining with a special vocabulary] would have been possible, perhaps, by long experiment, to arrive at a vocabulary of two or three thousand words adequate to define the whole number of words dealt with. [...] It seemed better to make definitions on the general principle (1) that common words should be explained by means of other common words (with the useful addition of synonyms which are less common) or by means of pictures or diagrams, and (2) that less common words (likely to be met with only when the learner has already acquired a vocabulary of several thousand words) should be defined by the use of a wider vocabulary. (ALD1, v)

As is clear from the above quotation, the editor admitted “less common” synonyms, and used them to complement analytical definitions; for example:

stingy 1 unwilling to spend money; mean in money matters; miserly (ISED)

The less common synonym *miserly* in the above definition should not pose a comprehension problem for the learner, provided that the user is familiar with the lexicographic convention that implies that the word has the same meaning as that indicated in the preceding definition. This strategy had a potential of expanding the learner’s vocabulary, and could be useful in production.

ISED extensively drew on COD3 (1934), a general-purpose dictionary extremely popular with the British audience throughout the first half of the century. Hornby adjusted the macrostructure of COD with the foreign language learner’s needs in mind. He took the majority of the entries from the source dictionary, and deleted those considered “unsuitable for foreign learners at pre-university level” (Cowie 1999a: 47). He also adjusted the COD definitions to meet the needs of FL learners. One strategy was to replace a potentially difficult word with the one that would be familiar to learners, while retaining the overall structure of the definition (Cowie 1999a, 47–48):

malevolent ... Desirous of evil to others (COD3)

malevolent ... wishing to do evil to others (ISED)

Another strategy was to unpack a noun phrase through a participial construction. Thanks to this strategy, a specialist term could be rendered into simple words, as dealt with in Cowie (1999a, 48):

malnutrition *n.* Insufficient nutrition (COD3)

malnutrition *n.* not getting enough food or the right sort of food. (ISED)

While many ISED definitions were simpler than those of COD3, a number of other definitions retained the original wording that was hardly easy for learners. In the

following definitions, the editor has adopted from *COD3* rare words such as *latent*, *pricking*, *perplexity*, *dilemma*, *stale*, and *down* used as a noun in the meaning of “feathery or fluffy material” (the last one under **powder-puff**). All of the words could present a comprehension problem for intermediate learners.

potential	Capable of coming into being or action, latent	(<i>COD3</i>)
potential	latent; undeveloped, but capable of coming or being brought into effective action	(<i>ISED</i>)
powder-puff	soft pad usu. of down for applying p. to skin	(<i>COD3</i>)
powder-puff	a small pad of down, etc., for applying powder to the face	(<i>ISED</i>)
pungent	affecting organs of smell or taste, or skin &c., with pricking sensation ...	(<i>COD3</i>)
pungent	(of smell or taste) causing a pricking sensation ...	(<i>ISED</i>)
quandary	A state of perplexity, difficult situation, practical dilemma	(<i>COD3</i>)
quandary	a state of doubt, perplexity, or uncertainty; a dilemma	(<i>ISED</i>)
rancid	Smelling or tasting like rank stale fat	(<i>COD3</i>)
rancid	with the smell or taste of stale, decaying fat or butter; not fresh	(<i>ISED</i>)

Other sources of difficulties could be idiomatic expressions, such as *man-of-war* in the definition of **powder-monkey**; and collocations with a word used in a non-literal meaning, for example, *be on friendly terms* in the definition of **hobnob**:

powder-monkey	(formerly) a boy employed on a man-of-war to carry powder to the guns	(<i>ISED</i> 1942)
hobnob	be on friendly terms	(<i>ISED</i> 1942)

3.3.1.3 Conclusion

NMED and *ISED* provided models of vocabulary control to be developed and perfected in subsequent decades. Both dictionaries occasionally admitted infrequent words and combinations of words that could be puzzling for the learner. Both dictionaries used extensively the defining practices inherited from native-speaker's dictionaries (cf. Rundell 1988, 130–132). Hornby's policy of using less-common synonyms (to accompany analytic phrases) improved the productive value of *ISED*, but the success of this strategy relied on the learner's familiarity with the lexicographic convention.

3.3.2 Revisions of ALD and emergence of new competitors: 1963–1986

Following the transformation of *ISED* into *ALD1* in 1948, the dictionary entered a period of unchallenged domination in the EFL market. The period lasted 30 years until 1978, when the first competitor emerged: *LDOCE1* (see also Rundell 2008a;

Stein 2002a, 102–3). Within this time, the Oxford dictionary was revised twice: in 1963 (*ALD2*) and 1974 (*OALD3*).

3.3.2.1 *ALD2 (1963)–OALD3 (1974)*

In *ALD2* (1963), there was no radical change in the defining policy. Although the editor did not use a special defining vocabulary, he attempted to make the definitions “as simple as possible” (*ALD2* vi). Cowie (1999a, 93) provides evidence for removal of certain difficult and formal words, such as *wretched* and *forsaken* in the definitions of *desolate* (*adj.*), but this type of revision was occasional rather than regular.

The addition of some rare items in *ALD2*, such as *ironmongery* in the definition of **hard-ware**, may be surprising but it fits the policy of supplementing analytical phrases with “less common” synonyms:

hard-ware articles made from metal, as pots and pans for cooking, locks, nails and screws, knives, axes, etc. (*ISED*)

hard-ware ironmongery; metal goods for domestic use, such as pans, nails, locks (*ALD2*)

Simplicity of definitions was not the overriding goal for the editor. Revisions towards greater simplicity had to be balanced by changes towards greater conciseness and accuracy of definitions (Cowie 1999a). An example is given below, where the new definition makes use of an abbreviation (*sb.*) and ellipsis in the second and the third clause:

accuse say that a person has done wrong, that he has broken the law, or that he is to be blamed (*ISED*)

accuse say that (*sb.*) has done wrong, broken the law, is to be blamed (*ALD2*)

As a result, the changes above made the definition shorter, though not necessarily more intelligible to the learner because of using *wrong* in the nominal sense.

In the definition of *hard water* (below), although a word *lime* has been replaced with a self-explaining *mineral salts*, another rare word *lathering* has been used to make the definition more precise:

hard water containing too much lime and therefore unsuitable for washing (*ISED*)

hard water containing mineral salts that interfere with the lathering of soap (*ALD2*)

Another change towards greater precision can be found in the definition of *powder*, which specifies different ways of obtaining the substance by using a series of verbs that are semantically narrower (“crushed, rubbed, or worn”) than the verb from the earlier definition (“reduced”). What may create an extra barrier to the

learner is a word *worn*, which, though adds to the precision of the definition, is used in a non-central meaning (of decomposing or damaging the substance):

powder solid material which has been reduced to very fine particles (ISED)
powder (kind of) substance that has been crushed, rubbed, or worn to dust (ALD2)

The following edition, *OALD3*, came out in 1974, eleven years after the publication of *ALD2*. According to Cowie, who joined on the editorial staff of *OALD3*, the revision of this edition was determined by some lexicographic problems that had been left unresolved in the previous edition, including the treatment of pronunciation, the verb-pattern system, the provision of illustrative examples, and the coverage of scientific and technical terms (Cowie 1999a, 97). To that end, Hornby organized a team of linguists from the University of Leeds, with Cowie being in charge of the treatment of phraseology (Cowie, 97).

As far as the vocabulary of definitions is concerned, it was not subjected to radical revision. Occasionally, the editor replaced rare words with more common ones. A good example is given below, where *solitary* has been changed for *lonely*:

solitude solitary place (ALD2)
solitude lonely place (OALD3)

On the other hand, many other rare and semantically more or less opaque lexemes have been either retained (e.g., *ebony*, *fir*, *goad*, *incite*, *pickle*, *rubble*, *teak*, *ironmongery*, *wearily*, *lather*) or added (*daffodil*).

Summing up, in the early editions of *OALD* intelligibility of definitions was not the only goal pursued by the editors. Revisions were also motivated by the desire to make them precise and concise.

3.3.2.2 LDOCE1 (1978)

LDOCE1 was the first serious competitor to *ALD*. As Cowie (2002, 106) remarks, the dictionary deserves a special place in the history of EFL lexicography as a true follower – and for many years the only one – of the tradition of defining with an explicitly restricted DV, which was initiated over thirty years earlier by *NMED* (1935). According to Paul Procter, who was the dictionary editor, the benefit of using a restricted vocabulary is “the fulfilment of one of the most basic lexicographic principles – *that is that the definitions are always written using simpler terms than the words they describe*” (*LDOCE1*, ix). The dictionary restricted the vocabulary of definitions and illustrative examples. The restrictions were imposed on the selection of simple and complex words, word categories, and the meanings of the words used.

LDOCE1 was the first large-scale dictionary compiled with a heavy use of computers (Cowie 1999a, 120; Fontenelle 2009). Unlike in earlier lexicographic projects, the *LDOCE* database system allowed the lexicographers to easily search and analyse various categories of the dictionary entries, including definitions. This was possible thanks to the marking of all components of the entries with numerical symbols, some of which encoded entry parts not intended for publication (Cowie 1999a, 120). Thanks to the complex mark-up, all the definitions could be gathered together and machine-checked. This ensured that the words used in definitions belonged to the restricted vocabulary.

It is worth noting that by that time, computers had already been employed in dictionary production. One of the first such dictionaries was the Random House Dictionary (*RHD*), compiled in 1966 under the editorship of Laurence Urdang (Cowie 1999a: 120; Landau 2009: 228). What was innovative at the time was Urdang's database system which recognized various entry elements. However, unlike in *LDOCE*, the system did not allow one to encode entry elements that were not intended for printing (Urdang 1966; Nesi 2009: 458). Moreover, the computer was used only in some stages of editing and production (Landau 2009: 228). The database was cumbersome to use, because the lexicographers had to communicate with the computer by means of a command line interface. Another dictionary worth mentioning is *OALD3* (1974). Compiled a few years before *LDOCE*, the database system used in *OALD3* was unable to recognize different entry elements, so it was impossible for the lexicographers to gather together a particular information type and perform a focused examination of its content (Cowie 1999a: 120).

LDOCE1 came with a list of 2215 (Xu 2012) defining vocabulary items appended to the back matter. A large part of the list was drawn from the *GSL* (1953). As Procter explains, the items

were selected by a thorough study of a number of frequency and pedagogic lists of English, particular reference having been made to A General Service List of English Words (Longman, 1953, reprinted 1977) by Michael West.

(*LDOCE1*, viii–ix)

The compilers of *LDOCE1* drew heavily upon the *GSL*, but did not copy it uncritically. The content of the latter had to be scrutinized in the light of changes in the language and the intended audience. It was clear that the *GSL* had become outdated and required adjustments to suit the needs of dictionary users and lexicographers (see 1.5.3). My comparison of the *LDOCE1* DV and the *GSL*, shows that they share approximately 1900 items, whereas the former has over 300 items (13%) that are not in the latter.⁸ According to Ayto (1984, 53), the vocabulary was

8. The size of the *GSL* was computed by including all the headwords and members of the families as provided in West's *General Service List of English Words* (1953). The count was based

modified by incorporating words that are up-to-date and particularly useful for definers. Nakane (1998, 47) observes that among the items added were scientific and technical terms (e.g., *acid, alcohol, atomic, breed, cell, chemical, chemistry, microscope, nylon, plastic, television, tyre*), body-part terms (e.g., *ankle, bowel, breast, chest, elbow, forehead, muscle*), linguistic terms (e.g., *adjective, adverbs, consonant, participle, phrase, singular*), names of weekdays and months, common names of plants and animals (*ant, banana, chicken, fox, lamb, mosquito, onion, potato, tiger*), and religious terms (e.g., *Bible, Christian, Hindu(ism), Jew(ish), Muslim*).

Procter followed the West convention of using a restricted DV. In addition, he allowed for the occasional use of words which were low on frequency lists but necessary in some definitions. They were normally printed in small capitals, and sometimes provided with an explanatory gloss:

gherkin a small green vegetable which is usually eaten after being PICKLED in VINEGAR (=kept in a sour liquid); type of CUCUMBER (*LDOCE1*)

In the case of derivatives from outside the list, only the bases were capitalized, for example, EMBARRASSing, EMBARRASSed, EMBARRASSes, in the definitions for **embarrassment**. In this way, the unknown base was distinguished from the known suffix. In turn, *removable* and other derivatives whose bases were on the DV were printed without special marking.

The DV included frequent and regular affixes, allowing the lexicographers to use them in definitions to form a wide range of derivatives not entered in the list. However, this strategy did not always result in transparent derivatives, for which the editor was taken to task (see further below). The list also contained a small set of multiword items, with one phrasal verb (*wrap up*), and several common expressions (especially relational ones) which might have been useful in defining (*so-called, upside-down, according to, no one, owing to*). Most entries were listed as single vocabulary items, except for several cases of inflected forms and derivatives which appeared in parentheses after the canonical form, as in *affair(s), alcohol(ic)*, or appeared as separate items, as in *actor, actress, humour, humourous* (Nakane 1998).

The restricted DV of *LDOCE1* came under considerable criticism. In an experiment conducted on the machine-readable version of this dictionary, Jansen et al. (1987) reveal a number of weak points of the DV. One of them is that the editors relied excessively on affixation in the belief that a learner would be able to understand compounds and derivatives. Among the problematic words were *freedom, business, forth*, and *unless*, which the lexicographers excluded from the

on word types, excluding part-of-speech homographs. Thus, *risk* n. and *risk* v. were counted as one item.

list but used in definitions on the assumption that the user would deduce their meaning from the structure. As Jansen et al. remark, the word *freedom* is essentially transparent in meaning as far as one of its senses is concerned, but the other words do not have compositional meanings, and thus can be more difficult for a learner to interpret. In other words, for a learner who knows the meaning of *-dom* (which is defined in *LDOCE1* as “the state of being something”), it is easy to conclude that *freedom* means “the state of being free”, which is the sum of the meanings of *free* and *-dom*. However, the meaning of *business* does not arise from the sum of the meanings of *busy* and *-ness*, nor do *forth* or *unless* invite straightforward interpretation.

A related issue has been raised by Herbst (1986), who maintains that the list had redundant elements in the form of self-explaining derivatives, such as *alcoholic*; *addition*, *activity*, which were provided along with their base words and suffixes. If one assumes that a DV should list word formations whose meaning cannot be deduced from their structure, the above derivatives are unnecessary (Herbst 1986, 114; Herbst 1996). Other superfluous elements have been found by Nakane (1998); for example, *although* and *though*, which figure as separate entries, are almost interchangeable in use, whereas *popular* is entered twice with no POS label (Nakane 1998).

Jansen et al. (1987) report that some items in the DV, in particular *amusing*, *anyone*, *rid*, and *someone*, were not given a prominent place in the dictionary, as they did not figure as entry headwords. The last word, *someone*, was very frequent in the dictionary definitions, in which it occurred 1016 times (see Fontenelle 2009, 420). However, the word was hidden in the entry for *somebody* as a variant (Fontenelle, 420), which practically made it impossible for the user to find the word. Note that the entry for *someone* would be introduced in *LDOCE2*.

Jansen et al. (1987) also report a discrepancy between the list and the actual words used in the definitions. For example, *business*, *forth*, *hole*, *unless*, and *whose* were used in definitions but were missing in the list. Some of them are function words, which are very frequent in the language as well as in the dictionary, for example, *whose* under **inscrutable** “... whose meaning is hidden or hard to find out”. As the researchers note, the inclusion of *business* and *whose* in the DV would have been justified by the fact they are recorded with a high frequency in West’s *General Service List of English Words* (1953). Note that according to the GSL, *business* has 946 occurrences in a count of 2.5 million running words, and *whose* – 722. All the above words but *forth* would appear on the list in *LDOCE6*.

Another point of criticism is the fact that the DV recorded only one phrasal verb (*wrap up*), implying that all the other occurrences of phrasal verbs in definitions should be treated as combinations of verbs plus adverbs/prepositions (Jansen et al. 1987). This is, however, not the case, for example, **be hard on** “to wear

(something) out easily or quickly” (*LDOCE1*). Other similar cases are mentioned by Herbst (1986), for example, *wear away* and *give up*, all being used in definitions on the grounds that the elements belong to the DV.

Finally, it was impossible to establish a part of speech for many DV items, because POS labels were not provided systematically in the list (Jansen et al. 1987). For example, *church* may have been a noun, an adjective, or a verb; and *left*, an adjective, noun, adverb, or past participle (Jansen et al. 1987). This may have raised problems not only for the lexicographers but also to the dictionary user, who could not be certain of what word category to expect in definitions (Fontenelle 2009, 420).

According to the editor, “a rigorous set of principles was established to ensure that only the most ‘central’ meanings of these 2,000 words, and only easily understood derivatives, were used.” (*LDOCE1*, ix). To what extent the above constraints were fulfilled by the *LDOCE1* lexicographers was examined by Jansen et al. (1987)⁹. The scholars identified several problems, which are discussed below.

The researchers found a number of ambiguities resulting from the use of polysemous words, phrasal verbs, and idiomatic expressions in definitions (Jansen et al. 1987, 1). Having examined a subfile of all the entries describing the DV items, they found that only 20.19% of the items were monosemous (Jansen et al. 1987, 84). This means that the great majority of the DV was to some degree ambiguous. Although this finding should come as no surprise as polysemy is a property of most high-frequency words it nevertheless ran counter to the policy of centrality of meaning declared by the editor. Furthermore, as the researchers argue, if they took into consideration only the first sense of each headword, which is what many learners do when interpreting the meaning of a word, the senses would account for only 40% of the definitions. In addition, they also found that the first sense not always provided the central meaning, which could result in learners’ failure to identify appropriate meanings of defining words.

In an attempt to verify the editors’ claim that the DV items were used only in their “most central meanings”, Jansen et al. estimated the degree of ambiguity of each defining word. This value represented “the number of possible senses and idioms” that was attributed to the entry for a defining word in *LDOCE* (Jansen et al. 1987, 84–85). In the case of verb entries, the researchers also took into account phrasal verbs. It was clear that polysemous words had a high degree of ambiguity. After calculating the values for each defining word, the researchers counted the frequency of occurrence of the headwords in the *LDOCE* definitions. They came to the conclusion that many defining words were used in non-central

9. For the criticism of this claim, see also Neubauer (1989).

meanings and that the degree of ambiguity of the words correlated highly with their frequency of occurrence in the definitions. For example, the noun *hand* was reported to have a degree of ambiguity 52, which accounted for 12 senses of the word and 40 idiomatic expressions recorded in the entry for **hand**. This value went hand in hand with the high occurrence of the noun (134 times) in definitions. Most of these occurrences (110) corresponded to sense 1, whereas 5 instances to sense 2, 4 to sense 4, and 6 to sense 6. In the study of the verb *run*, the researchers showed that while 74 occurrences referred to the “act of moving fast”, all other uses (19), were “to a certain extent idiomatic” (Jansen et al. 1987, 88). The latter group involved 16 cases of phrasal verbs.

Another source of ambiguity was the verb *break*, which was found to form, what Cowie calls, restricted collocations, that is, word combinations in which one element was used in a non-literal meaning.¹⁰ The verb was typically used in this meaning in *break the law*, *break the habit*, *break the rules*, and *break a record* (see also Cowie 1999a, 111). Phrasal verbs accounted for approximately 25% of the occurrences of *break* (Jansen et al. 1987, 89). It is not surprising that collocations and phrasal verbs were difficult to control because core words are particularly productive in forming such combinations of words (Beekman and Callow 1974).

In order to solve the problem of ambiguity of polysemous words, Jansen et al. (1987) suggested the use of superscript numbers to cross-refer the user from the word to its appropriate sense (Jansen et al. 1987, 84; cf Neubauer 1984, 120; Herbst 1986, 114). The lack of a superscript would mean that a word should be understood in the basic sense. In practice, however, this convention would not seem to make the interpretation of the definition immediately clear-cut to the learner. Furthermore, it would cause a significant increase in the number of the DV items (cf. Xu 2012, 372), which is clearly not the aim of vocabulary control.

Finally, some definitions in *LDOCE1* were criticized for being imprecise, a fault which is attributed to the necessity to meet the requirements of the DV policy (Wekker and Hyams 1979, 411). Imprecision was especially noticeable in definitions of technical and scientific terms (Stein 1979; Herbst 1986). For example, as

10. Cowie uses the term *restricted collocation* to refer to such word combinations that are distinct from free combinations and idioms (Nesselhauf 2005, 14–16). The distinction is made on the basis of two criteria: transparency and commutability. As for the former, restricted collocations have at least one element that is used in a non-literal meaning, for example *perform* in *perform a task*, but unlike idioms, the entire combination tends to be transparent. Unlike free combinations, such as *drink tea*, restricted collocations display only limited commutability, or substitutability, which means that their elements are substitutable to a limited extent. For example, we cannot say that the milk is rotten or that the egg is rancid (Fontenelle 1994, 2; see also Aisenstadt 1979).

Cowie observes, the definition of *lobster* did not mention *claws*, which are a distinctive feature of this animal, and it classified the word using the “awkward and misleading” *sea animal* rather than the biologically sound term *shellfish* (Cowie 2012, 112).

As can be seen, the *LDOCE1* policy of vocabulary control was praiseworthy but its systematic execution was difficult to attain. The limitations imposed on meanings and word-formation processes were particularly difficult to implement in a consistent way.

3.3.3 Conclusions

Following Neubauer’s classification (1989, 900), we can conclude that English dictionaries up to the 1980s adhered to three approaches to controlling a definition vocabulary:

- a. No or little control over the lexical content of definitions;
- b. Use of a simple vocabulary but without explicit restrictions;
- c. Use of a defining vocabulary selected from a specified list and, in addition, an occasional use of words outside the list.

Approach (a) is typical of native-speaker dictionaries, especially those which prioritize precision in defining. The early history of learners’ dictionaries shows that the more restrictive approach (c) precedes the more permissive one (b). The former, which relied on an explicitly restricted DV, was initiated by West in *NMED* but found no followers for over 40 years. It enjoyed a rebirth with the emergence of Procter’s *LDOCE* (1978). Until that year, EFL lexicography had been dominated by a more relaxed approach to vocabulary control (b), with no explicit restrictions. This approach was implemented by Hornby in *ISED* and continued in the early editions of *OALD*. However, both approaches (b) and (c), suffered from weaknesses, and they would be subjected to further revisions in the years 1987–2015.

3.4 Review of previous studies of defining vocabularies

Over thirty years ago Anthony Cowie made a remark that there had been “[v]ery little research” into “the reactions of foreign students to the simplified definitions of *LDOCE* as compared with corresponding definitions in a non-controlled vocabulary” (Cowie 1989, 47; see also Cowie 1990, 372). This observation is still valid today not only in the context of *LDOCE* but also other learners’ dictionaries. In spite of a short history of research into dictionary use (Nesi 2014), various aspects of dictionary use have been extensively studied in recent decades (Tomaszczyk

1979; Tono 1989; Bogaards and van der Kloot 2001, 2002; Lew 2002, 2004; Nesi and Haill 2002; Chi 2003; Dziemianko 2004, 2012b; Lew and Dziemianko 2006).¹¹ Nevertheless, there is little empirical research on the perception and comprehension of definitions written in a controlled vocabulary vs. non-controlled vocabulary. The previous studies differ in methodology and scope; most of them address research questions other than vocabulary control, such as the effectiveness of sentence definitions or entire entries (for a survey of dictionary user studies see Nesi 2014), and those that explore this topic narrow the focus on the early EFL dictionaries published up to 1995. This is surprising, given the fact that the period following the year 1995 was prolific in the production of learners' dictionaries. It saw not only a series of revisions of *LDOCE* (2003, 2009, 2014), *OALD* (2000, 2005, 2010), *COBUILD* (2001, 2003, 2006, 2010), *CALD* (2001, 2008, 2013), and the emergence of *MEDAL* (2002, 2007), and *MWALED* (2008). Little research has been done on the lexical content of various defining vocabularies. To my knowledge, there are virtually no studies estimating the feasibility of defining vocabularies along the research lines proposed by Laufer (1989), Nation (2006), Webb and Nation (2008), and Nation and Webb (2011). One notable exception is Xu (2012), who applied this methodology to the *LDOCE* DV lists (up to *LDOCE*5). Another field that is virtually unexplored is that of vocabulary control in English dictionaries for native speakers, both in terms of the recent history and potential usefulness for less competent users of the language. What follows is a review of user-oriented studies of vocabulary control, and the studies of the content of defining vocabularies.

MacFarquhar and Richards' paper (1983) is seminal in addressing the perception of dictionary entries by foreign learners. The researchers asked students to compare 60 corresponding entries from *LDOCE*1, *OALD*3, and the native-speaker *WNWD* (1972), and then "indicate which definition they found easiest to understand" (1983, 119). Each dictionary made use of a different approach to vocabulary control: *LDOCE*1 drew upon an explicitly limited set of defining vocabulary, *OALD*3 simplified definitions with no recourse to such a vocabulary, while the native-speaker *WNWD* made "the least effort to limit its definitions to simple words" (MacFarquhar and Richards 1983, 115). Each student was provided with a card with the corresponding complete entries side by side. Worth noting is the fact that the order of entries was randomized and no indication of the source was provided so as to reduce the risk that the subjects' choice of the entry could be influenced by their awareness of the source dictionary. The researchers report that learners preferred the *LDOCE*1 definitions. Interestingly, while comprehensibility of definitions from the two learners' dictionaries

11. For a survey of empirical research into dictionary use, see Nesi (2014)

was judged higher than that of the dictionary for native speakers, the difference between *LDOCE1* and *OALD3* was much greater than that between *OALD3* and *WNWD*. This finding suggests that definitions written in an explicitly limited defining vocabulary are perceived as more intelligible than definitions written without such restrictions (see also MacFarquhar 1985). I will use this finding as a basis for hypothesis H7 (see 3.4.2).

Yet the methodology of MacFarquhar and Richards' study has been criticized for insufficient control of extraneous variables. Although the researchers deserve credit for taking care of reducing the effect of word order on the subjects' decisions, they failed to control the effect of other information types located in various parts of the entries, for example, complex verb patterns in *OALD3*, abbreviated etymologies in *WNWD*, and the tilde, which was more likely to be found in *WNWD* and *OALD3* than in *LDOCE1* (Nesi 2000, 47). Although all these factors are irrelevant to the defining practice, they may have affected the subjects' perception of the definitions. Hence, there is a risk that subjects made decisions about intelligibility of definitions by comparing entire entries. As Nesi (47) remarks, "[w]hat MacFarquhar and Richards were really monitoring was the users overall impression of the *look* of the entry". Although this study does not prove what the researchers had set out to demonstrate, namely "practical consequences of different methods of definition" (MacFarquhar and Richards 1983, 119), it unequivocally shows that foreign learners prefer entries in *LDOCE1* to entries in *OALD3* or *WNWD* (Nesi 2000, 47).

Three years later in 1986, Herbst conducted a research very similar to that of MacFarquhar and Richards. However, unlike the previous researchers, Herbst took just definitions, with no examples or grammatical information, as the object of his study. He collected definitions of 14 words from three learners' dictionaries: *LDOCE1*, *OALD3*, and *CULD*, and presented them to foreign language students. Marked with letters A, B, and C, the definitions were sometimes given in a different order. Of all these dictionaries, only *LDOCE1* made use of an explicitly restricted vocabulary; the others followed a less rigorous approach without specifying the content of the defining vocabulary. The subjects had to rate the definitions according to their comprehensibility and adequacy. The latter criterion apparently referred to the accuracy of the definitions as judged by the learners. The students were asked to indicate definitions which they found easiest to understand, most difficult to understand, most adequate, and least adequate. Herbst's subjects demonstrated a marked preference for the *LDOCE1* definitions, rating them highest with respect to both criteria. The *LDOCE1* definitions accounted for 50.3% of all the definitions that were judged easiest to understand, and for 66.9% of all the definitions deemed most adequate. The researcher attributed these findings to the *LDOCE1* defining policy, though the sample was small (14 words and

22 subjects) and no far-reaching generalizations were possible. When discussing this policy, Herbst observes that in spite of lexical restrictions,

there is no limitation to the words actually used in definitions since it is an essential part of the LDOCE defining policy that words outside the defining vocabulary can be included in definitions where it seems appropriate

(Herbst 1986, 112)

He notes that it is up to experienced lexicographers to decide where a non-DV word should be used, and where it is necessary and appropriate. Herbst (112) poses the question of whether equally comprehensible definitions can be achieved by “merely instructing them [i.e. lexicographers] to use the most simple words possible”. This question calls for a closer inspection of different approaches to defining vocabulary, with and without an explicit vocabulary list.

The above studies focused on how definitions written with a controlled vocabulary are perceived by learners, not on how useful or effective they actually are in real language situations. A few researchers have compared the effectiveness of different definitions, though without special focus on a controlled DV. This research will be presented below.

Cumming, Cropp and Sussex (1994) examined ESL students’ preference for and effectiveness of two definition formats in comprehension and production tasks. They directed their attention to sentence definitions of *COBUILD1* in contrast to traditional phrasal definitions of *LDOCE2*, and tested them in different conditions, either with or without additional usage examples. They found no significant differences in the effectiveness between the definition formats, and between formats with and without additional context (Cumming et al. 1994, 375). Worth noting is that the study was designed to test not only students’ understanding of definitions (and illustrative examples) but also their ability to distinguish between correct and incorrect use of the unfamiliar words. Therefore, the entries to which the subjects were exposed contained not just definitions but also the prepositional information and the part-of-speech of the words tested; for example: “*quibble* (about, over, with) verb: to argue about small unimportant points or details” (Cumming et al. 1994, 372). What is interesting in Cumming et al.’s study, although the subjects preferred the *COBUILD*’s sentence definitions, there was no or little correlation between their preference and various performance measures, in both production and reception. This finding revealed limitations of MacFarquhar and Richards’ study, showing that students’ preference may not necessarily correlate with the task performance (Nesi 2000, 49). The high level of users’ satisfaction from *COBUILD* was also reported by Harvey and Yuill (1997). Although Cumming et al. were interested in the effectiveness of two definition formats, one cannot rule out the possibility that the subjects ratings were influenced by different

levels of difficulty of the defining vocabulary, if any difference existed, and perhaps by the approach to the selection of this vocabulary.

Drawing on the previous research techniques, Nesi (1998) investigated how meaning was conveyed in a set of dictionaries designed for foreign learners (*OALD5*, *LDOCE3*, *COBUILD2*, *CIDE*) and a dictionary for native speakers (*COD8*). Nesi selected five nouns designating household objects (*colander*, *insole*, *plunger*, *shoehorn*, *spout*) and extracted their definitions and illustrative examples (the latter when given). She administered a three-stage task to 158 subjects, all being foreign learners. Having checked the learners' familiarity with the words, she asked them to read the entries and judge them in terms of their perceived helpfulness. Following earlier studies, the subjects gave an overall assessment of helpfulness of entries, without specifying details of their use. Then the subjects were asked to write their own sentences containing each of the words, and identify the pictures that best illustrated the words. In the perception task, the subjects found *LDOCE3* and *CIDE* most helpful, and the *COD*, least helpful. However, neither the production task nor the identification task showed a statistically significant difference between the observed and expected frequencies of "right" (or "wrong") answers.¹² The results prevented the researcher from drawing firm conclusions about the superiority of one dictionary over another with respect to the helpfulness in production or picture identification tasks. As in Cumming et al.'s study, the learners' perception of entries did not correlate with their helpfulness.

Two other studies are worth mentioning as they address the issue of potential usefulness of definitions in native-speaker dictionaries for foreign learners and definitions in EFL dictionaries for native speakers. McCreary and Dolezal's (1999) research was designed to determine the usefulness of a native-speaker dictionary (the *American Heritage Dictionary*, 1982, henceforth *AHD*) for ESL students in a comprehension test. They devised a multiple choice test to examine student's understanding of selected words (e.g., *juxtaposition*, *shrine*, *sheen*) under differing conditions: with the aid of the dictionary only, or the context only, or the combination of the dictionary and the context. They find that the subjects who were exposed to the *AHD* entries plus the context scored higher than the other groups of students. What is important is that the students who used the dictionary only did not perform better than the students who used the context only. The study demonstrates the importance of context for comprehension of difficult words. It also shows that the *AHD* definitions are difficult for the ESL students. Although the researchers do not discuss the reasons for this difficulty (e.g., the unrestricted

12. The difference was calculated with a chi-square test.

vocabulary of definitions), they conclude that the “ESL learner may indeed be better off with a learners’ dictionary” (McCreary and Dolezal 1999, 134).

In a later study, McCreary and Amacker (2006) set out to examine an opposite situation to that described above. They wanted to determine whether the entries in an EFL dictionary meet the receptive needs of native speakers of English. The dictionary under study was *MEDAL*. The researchers prepared a comprehension task in the form of a multiple-choice test for the same selection of hard words as in the earlier research (i.e. McCreary and Dolezal 1999). The test shows that American college students using *MEDAL* performed slightly better than the students who had the *Collegiate Dictionary* (11th edition) at hand, even though the former dictionary was not the type of work intended for the subjects. According to the researchers, the higher number of errors made by the *Collegiate* users should be attributed to the low-frequency words employed in this dictionary (McCreary and Amacker 2006). However, given the fact that students were exposed to entire entries, one cannot rule out the effect of other factors, such as the user’s inability to locate the appropriate sense, etc. Also the researchers were careful to make sure that the subjects really used dictionaries by asking them to underline the line or phrase that was most helpful in doing the test. This research lends support to McKeown’s belief that “it might be appropriate to develop a ‘learner’s dictionary’ for native speakers” (1993, 29).

Finally, let us look at Nesi and Meara’s research (1994) into the productive use of *LDOCE*, *OALD*, and *COBUILD*. The researchers do not specify which editions they used, but judging by the publication year of the paper, they must have been the 1980s editions: *LDOCE2* (1987), *OALD4* (1989), and *COBUILD1* (1987). 52 learners were assigned to one of three groups, each with access to entries drawn from one of these dictionaries. The subjects were asked to read the definitions of unfamiliar words and write sentences using these words. The researchers analyzed the sentences and identified three categories of errors:

- a. Failure to use the unfamiliar word (no target word in the sentence),
- b. Semantic error,
- c. and usage error (e.g., problems with transitivity, countability, or morphology).

The researchers found that the error categories (a) and (c) were relatively rare, and that there were no significant differences in the frequency of errors between the groups of dictionary users. They noted that the majority of errors were of semantic type (b), and were presumably caused by misunderstanding of the definitions and neglect of collocation or connotation, *inter alia*. This type of error occurred significantly more frequently in the sentences produced by the *OALD4* users than by the *COBUILD1* users, which may have been due to “the special characteristics of

the *OALD* defining style” (Nesi and Meara 1994, 8). It should be added that both of the dictionaries used non-controlled definitions, and that the former dictionary adhered to the traditional defining style while the latter used full-sentence definitions. Interestingly, the subjects who had access to the restricted definitions of *LDOCE2* made nearly the same number of semantic errors as the *COBUILD1* users. It is likely that the use of the restricted DV in *LDOCE2* contributed to their users’ success.

Nesi and Meara identified a number of errors which were attributed to the so called *kidrule* strategy described earlier by Miller and Gildea (1987) in a study of dictionary use by children. The strategy involved several stages, including reading a definition, selecting a familiar part of the definition, constructing a sentence with this part, and finally replacing the part with the word being defined (Nesi and Meara 1994). For example, a subject who was familiar with the last word of a definition of *tenet* “opinion, belief, principle, or doctrine held as true” composed a sentence *The news was very tenet*, using *tenet* in place of *true*. Thus, Nesi and Meara observed that even adult non-native users of EFL dictionaries have a tendency to misinterpret definitions by over-relying on definition words that do not carry the essence of meaning. Although in the present study I did not test production skills of subjects, I used the *kidrule* strategy to formulate some responses for the multiple choice test (see Chapter 6).

Aside from user-oriented studies, a few researchers have conducted empirical studies of the content of a restricted defining vocabulary, notably in *LDOCE*. Jansen et al.’s study (1987) is a computer-aided verification of the editor’s claims regarding restrictions imposed on the defining vocabulary of *LDOCE1*. Using the dictionary database, the researchers found a number of weak points in the selection and presentation of the controlled vocabulary. For example, they note that the list does not include certain words that are actually used in the definitions and examples, some words do not even appear in the dictionary as headwords, and defining words are not used in their central meanings (for the results of Jansen et al.’s study see Section 3.2.2.2.).

The defining policy of *LDOCE1* has also been investigated by Herbst (1986). A comparison of 200 corresponding entries from *LDOCE1* and *OALD3* demonstrates that in the latter dictionary “four to five times as many words outside the *LDOCE* defining vocabulary are used than in *LDOCE*” (Herbst 1986, 108). This means that *OALD3*, which adheres to a less restrictive policy than *LDOCE1*, uses a far larger defining vocabulary. Herbst argues that many of these words are likely to present problems to the foreign learner, which is partly supported by the aforementioned test that the researcher has conducted on students’ perception of definitions of 14 words. A point that Herbst emphasizes is that the use of a restricted defining vocabulary combined with “the flexible policy” of admitting other words to

definitions imposes a lexicographical discipline which is of paramount importance to making definitions easily comprehensible for the foreign learner (Herbst 1986, 101). The researcher formulates principles for constructing a defining vocabulary.

Another researcher who paid close attention to *LDOCE* was Xu (2012). In his paper, he compared the defining vocabularies of the first five editions of the dictionary in terms of size, frequency, range (i.e. the occurrence of a word across sections of a text corpus), parts of speech, senses, affixes, and multiword expressions (Xu 2012). To evaluate the vocabulary of the *LDOCE* lists, he used the RANGE program created by Heatley et al. and the BNC word-family lists (2002). The researcher determined the coverage of definitions by the three BNC lists: the most frequent 1,000 word families, the second 1,000 word families, and the third 1,000 word families. He noted that more than 85% of the defining items in the first five editions of *LDOCE* are extremely frequent, as they belong to the first 3,000 word families, and thus should be familiar to the dictionary user (Xu 2012). Xu also identified definition-specific items that figure on the *LDOCE5* list but are not present in the first three BNC lists. In this way, he determined groups of words used in the definitions purely for defining purposes. When analyzing the *LDOCE* lists with respect to the inclusion of affixes, Xu made references to the “Oxford 3000”, that is the DVs of *OALD7* and *OALD8*, and the DV of *MEDAL*. Xu’s findings can be treated as a starting point for a further inter-dictionary comparison, which is undertaken in the present study. The current study takes into account the most recent edition, *LDOCE6* (2014), which was not covered by Xu’s paper.

Fox (1989) offers an illuminating analysis of words used in definitions of *Collins Cobuild Essential English Dictionary (CCEED)* (1988), a dictionary which comes with a list of the defining vocabulary in the back matter. The author compares the list with two other vocabularies: the frequency list obtained from the COBUILD corpus and the list of the controlled vocabulary of *LDOCE1*. She identifies a core vocabulary of 3,000 words used most frequently in the definitions of *CCEED*. The criterion for inclusion in the core vocabulary is that a word should occur 10 times or more in the definitions. Fox finds out that, while the majority of the defining words (over 2000) belong to the most frequent words in everyday English, there are differences in frequency ranks. She identifies several groups of vocabulary items that are more important (because they are used more frequently) in dictionary explanations than in everyday language; for example, superordinate nouns (*tool, vegetable, vehicle*), grammatical words (*grammar, paragraph, noun, phrase*), words indicating the function or register of particular words (*rude, polite, disapproval*), intensifiers used for “teasing out the subtle meanings of words” (*calmly, loudly, quietly*) (Fox 1989, 162–3).

As for the comparison of the *CCEED* defining vocabulary with the controlled vocabulary of *LDOCE1*, Fox deplores the indeterminateness of the Longman

defining vocabulary and admits that such a comparison is difficult to make on the basis of the list appended to *LDOCE1*. As she remarks, “it is not possible for a user to know whether a word that is in the list is used frequently, infrequently, or even not at all” (Fox 1989, 170). She observes that numerous superordinate terms that are not recorded in the controlled vocabulary of *LDOCE1* are used in *CCEED*, for example, *design*, *diagram*, *device*, *characteristic*. Among them are the words that convey pragmatic information in *CCEED* but appear to be redundant in *LDOCE1*, for example, *definitely*, *deliberately*, *disapproval*, *discuss*, *discussion*, *emphasize*, *describe*.

Being an editorial director of *COBUILD2*, Fox (1989) expresses her doubts about the validity of the method of restricting a defining vocabulary to an arbitrary fixed number of items and imposing them on lexicographers before they start writing definitions. In this approach, which has been adopted by *CCEED* and *LDOCE* but not *COBUILD*, there is no certainty that the items listed at the back of the dictionary will be known to the user (Fox 1989). Another problem is that while a restricted vocabulary makes a definition simple, simplicity is sometimes achieved at the expense of accuracy. Fox goes on to point out that, although a careful selection of a defining vocabulary is “of vital importance”, it is very problematic to predict the defining words in advance (Fox 1989, 154). The problem is that the vocabulary prepared “either out of a vacuum or out of a general frequency list” may contain the words that are never needed in definitions, and may lack the words that are very frequently needed (Fox 1989, 157). Therefore, the approach adopted in *COBUILD1* and subsequent editions (personal communication with Collins Dictionaries) is radically different from that of *LDOCE*:

We must use in [dictionaries] language that learners will understand. We therefore ask our editors to monitor carefully the language they choose to use. We ask them, wherever possible, to explain words in other words that are more frequent than the ones they are defining. This they can usually do, because we have computer statistics that tell us the relative frequency of words. It is possible to group them, and so to define one word in terms of a group of words more frequent than it is. (Fox 1989, 156)

Critical voices of a restricted defining vocabulary have been raised by other researchers. Stein (1979, 6) observes that “by using only a very restricted basic vocabulary for the explanation of lexical items, the explanations can become oversimplified, the definitions imprecise and vague, with constituent features missed.” A similar concern is given by Lew (2010, 293), who points out that by restricting the choice of words, one may lose “some of the discriminating power to tease out the finer shades of meaning.” Hanks (2009, 307) maintains that one of the dangers of this approach is “distortion of meaning” caused by the necessity to “fit the terms

available.” Carter (2012) claims that a vocabulary control may lead to a “clumsy or unnaturally circumlocutory” communication. This is the type of communication that never occurs in natural language, even though it makes use of well-formed grammatical patterns (see Sinclair 1984; Fox 1989). What is more, Carter (2012) argues that the unnaturalness of a definition may arise when a word outside the DV needs to be used, but is not, and is replaced by a word from the list. This situation may lead to a collocation issue. The same problem is mentioned by Lew (2013, 298), who stresses that lexical restrictions imposed on lexicographers limit their choice of collocations. Lew goes on to explain that “[t]he artificiality extends to unnatural collocational patterns, as the natural collocates may not be in the defining vocabulary set.” In the same vein, Kirkpatrick (1985, 10) explains that, “The worst feature of the defining vocabulary system is the constraints that it places on the definer.” She argues that there is a risk that this system will result in unnatural language “encouraging the reader to emulate this.” (Kirkpatrick, 10). Furthermore, controlled definitions are criticized for the fact that they may hamper the learner’s productive skills and the need to enlarge their vocabulary (Allen 1996). Learners have limited opportunities to learn new words from the definitions they read, “since the progression is always downwards” (Allen 1996, 47). For example, as Béjoint (2000, 69) explains,

[i]f *cat* is defined as ‘mammal that ...’, then the users can take the word *mammal*, look it up, etc., thus enlarging their vocabulary and their knowledge of the world; but if *cat* is defined as ‘animal that ...’, the chain is, so to speak, ‘short-circuited’.

The above criticism calls into question the sense of restricted DVs. Given numerous problems associated with them, Hanks (2009, 307) suggests that “[i]t makes better sense to insist that the wording and structure of definitions should be as simple as possible but as complex as necessary”. The researcher seems to call for a compromise between lexical restrictions and precision of expression.

On the other hand, there is an opposing school of thought. A strong argument in favor of using a restricted defining vocabulary is that it makes definitions accessible to less proficient learners by eschewing rare and other difficult words. According to Summers (1988), a limited DV gives learners a better chance of understanding definitions and assimilating new words for future use. Furthermore, as Rundell (2008a, 319) observes, the rigor of such a DV often encourages lexicographers to “look beyond obvious synonyms in order to get at the deeper semantic core of a word or meaning”. According to Quirk, “the strict use of the defining vocabulary has in many cases in fresh and revealing semantic analysis” (cited in Rundell 2008a, 319). Such analysis brings out the aspects of word meaning that often cannot be captured by synonyms or L1 equivalents. However, it may sometimes be challenging for lexicographers. As Rundell (2008a, 319) argues, the

problems associated with a restricted DV are often caused by “its inexpert application”, for example, when using a frequent word in a rare meaning (e.g., **disreputable** “having a bad name” in *LDOCE1*). Finally, as mentioned earlier, the study by MacFarquhar and Richards (1983) shows that learners prefer definitions that explain difficult words in terms of the words they already know. This finding is an important contribution to the issue of vocabulary control because it emphasizes the learner’s role as the final arbiter of such dictionaries (Summers 1988, 12).

There are researchers who advocate bilingual learners’ dictionaries whenever monolingual dictionaries fail to meet the learner’s needs. They stress the complementary role of these two dictionary types (Piotrowski 2001; Lew 2004; Adamska-Sałaciak 2010b). Bilingualized dictionaries are seen as a solution to the deficiencies of monolingual (and bilingual) dictionaries. Such dictionaries combine the elements of the two dictionaries, using the learner’s mother tongue (L1) and foreign language (L2) in the description of meaning. This type of dictionary was postulated a few decades ago by Atkins (1985, 22). Among possible improvements of a monolingual dictionary, she suggests that the dictionary be customized by inserting L1 equivalents in place of L2 signposts that normally introduce senses. She also proposes that L2 definitions of fixed phrases be supplemented by L1 equivalents of the phrases. There is also a possibility to write entire explanations in learners’ L1, as proposed by Ščerba (1995; see also Adamska-Sałaciak 2014). It is believed that all these strategies improve the explanatory power of monolingual dictionaries by introducing selected elements of bilingual dictionaries. A great advantage of using the learner’s mother tongue is that the user is instantly provided with a familiar anchor point, and does not lose time wading through definitions. This is especially important in receptive tasks (Hartmann 1993, 160; Dziemianko 2012a, 45; Adamska-Sałaciak 2014). In the Introduction to *ISED* (1942, v.), Hornby recognized that a foreign learner who refers to *lobster* “usually needs only to identify the new word”. Cowie (1998b, 258) argues that for numerous technical and scientific terms, an L1 equivalent can provide the learner with a quick and sufficient explanation.

The incorporation of bilingual dictionaries features is seen as a step away from the global dictionary towards the local one (Kernerman 2000, 828–829). This would be a hybrid dictionary which is designed to meet the needs of learners from the same linguistic background. It is believed that, if they do not solve all the problems resulting from restricted DVs, such as meaning distortion and unnatural and convoluted wording, they will help a foreign learner comprehend the meaning of a word being looked up and will reduce the consultation time.

It is worth noting that hybrid, or bilingualized, dictionaries have already been produced for learners. The first bilingualized dictionary, which appeared in print in 1966, was intended for the Chinese audience and was based on *OALD2* (Cowie

1999a, 192). Since then a number of other bilingualized dictionaries have become available to learners from various countries; for example, in the 1990s such dictionaries were distributed in over 30 countries (Kernerman 2000, 828; Dziemi-anko 2012a, 45).

Cowie's *English Dictionaries for Foreign Learners: A History* remains one of the most comprehensive accounts of the development of EFL lexicography from the inception to the late 1980s. Cowie surveys the vocabulary control movement and examines various aspects of the design of the major learners' dictionaries. The dictionaries are evaluated in chronological order. In a three-page section devoted to the DV in *LDOCE1* (1978), Cowie reviews the results of experiments by aforementioned Jansen et al. (1987), MacFarquhar and Richards (1983), and Herbst (1986). This analysis is continued in a section on *LDOCE2* (1987), where the author investigates additions and deletions to the part of the DV under letter C. Regrettably, Cowie's account of the history terminates at 1989, leaving major dictionaries published in the 1990s unexplored (for a critical review see Nichols 2001 and Kirkness 2001).

The gap in the research – as far as the defining vocabulary is concerned – was partially filled by Nakane (1998). Nakane compares the DV lists of three dictionaries published in 1995: *LDOCE3*, *OALD5*, and *CALD1*. The researcher identifies a core defining vocabulary of 1541 items common to the DV lists. She recommends it as a basis for English textbooks for students of junior and senior high schools in Japan.

Stein (2002b) uses a similar methodology and the same dictionaries to arrive at a core defining vocabulary. She emphasizes the role of paraphrasing in learner's vocabulary development. The researcher believes that learning and teaching can be improved by developing the ability to paraphrase (51). Because the DVs of the dictionaries have proved useful in paraphrasing a large number of lexical units which the dictionaries cover, she proposes a common core of these vocabularies as "the most useful lexical minimum" for learners at the intermediate level (52). Her Common Core Vocabulary (CCV) is intended to be used by learners as a basis for expressing a wide range of ideas. It consists of 2,139 lexical items (see also 2.1.7), and is larger than Nakane's because it distinguishes between part-of-speech homographs. For example, *act* appears twice on Stein's list, as a noun and a verb, but only once on Nakane's. The researchers are among the few who have determined a core vocabulary on the basis of DVs in dictionaries.

Michiels and Noël (1984) investigated the alleged syntactic complexity of *LDOCE1*, presumably caused by the introduction of the limited defining vocabulary. Using the electronic database of *LDOCE1* and of the *New Collins Concise English Dictionary* (CCED), which does not control the defining vocabulary, they analyze syntactic patterns expressing the "instrument-process" link (e.g., *a tool*

used in, a device made for, an instrument made for). The researchers do not find major differences in the use of the patterns, though they find in *LDOCE1* occasional instances of awkward definitions resulting from the frequent use of *apparatus* as a genus term. They note that patterns *used for*, *used to*, *used in*, *used as* are highly productive in both dictionaries, and that *LDOCE1* chiefly work with *apparatus*, *instrument*, *machine*, and *tool*, while *CCED* makes use of all the same words plus *device*. Thus, as far as the above patterns are concerned, the syntax of the *LDOCE1* definitions has not been impoverished by the introduction of the restricted defining vocabulary.

The purported syntactic complexity of restricted definitions was taken up by Kamiński (2020). He compared definitions in *OALD4* and *OALD5* with respect to definition length and the frequency of occurrence of 7 syntactic constructions that could make definitions difficult to read for less proficient learners. They included passive constructions (e.g., *is taken*), nominal phrases with participles (e.g., *unit composed*, *people fighting*), and reduced relative clauses following a complex noun phrase (e.g., *used in* “heavy axe with a long handle, used as a weapon”). The study showed that the introduction of the restricted DV in *OALD5* had no impact on the majority of the constructions. The only pattern which increased significantly in number was a nominal phrase with a postmodifying past participle (e.g., *unit composed*). The researcher also reported significantly longer definitions in *OALD5*, but concluded that there was no sufficient evidence to claim that restricted definitions are more difficult for advanced learners to read than definitions written with no such restrictions.

There are other papers that devote a great deal of space to defining vocabularies. Although some of them are not, strictly speaking, empirical studies, they provide valuable comments on the lexicographic practices regarding a DV. As a member of the *LDOCE1* team of lexicographers, Whitcut (1988) explains the principles underlying the construction of the explicitly restricted vocabulary in this dictionary and draws attention to alternative approaches. She discusses the strengths and weaknesses of the *LDOCE* approach from the perspective of the lexicographer and the user. One of the problems mentioned is a mismatch between accuracy and comprehensibility, which is illustrated by a definition of *lava* and the lexicographer’s dilemma: “should we define *lava* as the stuff that flows out of a ‘burning mountain’ or out of a *volcano*, which is the right word but falls outside our defining vocabulary?” (Whitcut 1988, 53). Such examples are many; for instance, a definition of *malaria* would not make much sense without mentioning *mosquito* (Michael Rundell: personal communication). In spite of the above disadvantages, Whitcut (1986, 118) believes that “comprehensibility is best achieved by [...] using an explicitly listed defining vocabulary for foreign learners.” This opinion stands in sharp contrast to that of Fox (1989: 156) mentioned earlier.

For an overview of the methods of controlling a defining vocabulary, the interested reader may turn to Neubauer's paper (1989). The researcher discusses a selection of definitions written within and without a controlled vocabulary and cites lexicographers' diverging opinions about approaches to vocabulary control. In Neubauer (1987), the researcher offers many useful tips for compiling a DV for a German dictionary. Enlightening analyses and practical applications of DVs are presented in Bogaards (1996, 2003, 2008), Stein (1979, 2002b), Sobkowiak and Kuczyński (2002), Bullon and Leech (2007), Fontenelle (2009), De Schryver and Prinsloo (2011), and Hiles (2014). The design of controlled vocabulary lists for ESL/EFL teaching and learning is presented in Chapters 1 and 2.

3.4.1 Summary of previous studies and implications for the current research

From the above survey of user-research, the following methodological points emerge:

1. Previous research on perception and understanding of definitions/entries by EFL learners has been conducted on the dictionaries published up to 1995 (*LDOCE1-LDOCE3*, *OALD3-OALD5*, *COBUILD1-COBUILD2*, *CULD*, *CIDE*, *WNWD*, *COD8*, *AHD*). This line of research is worthy of continuation on dictionaries published after 1995.
2. In the majority of previous studies, subjects have been exposed to either entire entries (MacFarquhar and Richards 1983; McCreary and Dolezal 1999; McCreary and Amacker 2006) or entry parts containing definitions plus other information types, such as illustrative examples and/or grammatical information (e.g., Cumming et al. 1994, Nesi and Meara 1994; Nesi 1998). The reason is that the researchers have been interested in examining the microstructural units larger than definitions, and the way they were helpful in both reception and production tasks (Cumming et al. 1994, 375; Nesi and Meara 1994; Nesi 1998). Herbst (1986) stands out from the other researchers in that he has collected information on learners' preferences exclusively through definitions (*LDOCE1* vs. *OALD3* vs. *CULD*) rather than entire entries (MacFarquhar and Richards 1983), and his primary interest was in a restricted defining vocabulary. This line of research deserves more attention.
3. The studies by McCreary and Dolezal (1999) and McCreary and Amacker (2006) are among the few which investigate how EFL learners and native English-speakers can benefit from dictionaries intended for different audiences: native speakers and learners.
4. All the studies presented above are concerned with a comparison of actual dictionary definitions rather than invented definitions. The advantage of the

latter study would be the elimination of the effect of external variables (Dziemianko 2012c). This is what the current study intends to do.

Summarizing the findings from previous research on the dictionaries published up to 1995, one needs to emphasize the following:

1. With respect to the research into the lexical content of defining vocabularies, it must be concluded that, of all the EFL dictionaries, *LDOCE* (up to *LDOCE5*) has received the greatest attention from scholars. Researchers have identified weaknesses of the *LDOCE1* policy of vocabulary control and suggested a number of improvements (e.g., Herbst 1986; Jansen et al. 1987; Xu 2012). When analyzing the vocabulary lists of *LDOCE1–LDOCE5*, Xu draws attention to an essential aspect of composition of such lists, namely how much vocabulary in the DV lists is covered by the words that should be familiar to non-native learners of English. The researcher uses BNC word family lists as a reference point against which to measure the vocabulary coverage of the lists. His analysis is confined to the DV lists in the first five editions of *LDOCE*.
2. There are few studies that determine a core defining vocabulary on the basis of dictionary definitions rather than a corpus of general language. Nakane (1998) and Stein (2002b) derive their core defining vocabularies from the intersection of the DVs of a few dictionaries published in 1995. This line of research deserves continued study.
3. As for user-oriented research, several researchers report learners' preference for definitions in *LDOCE* (MacFarquhar and Richards 1983; Herbst 1986, Nesi 1998) and *CIDE* (Nesi 1998) over definitions in other dictionaries including those for learners and for native speakers. Some researchers (MacFarquhar and Richards 1983; Herbst 1986) attribute this preference to the restricted defining vocabulary.
4. On the other hand, Cumming et al. (1994) demonstrate that students prefer definitions from *COBUILD1* to those from *LDOCE2*. The researchers explain that the reason is the different format of definitions: sentence vs. phrase. This finding is worth further studies.
5. Examination of learners' perception of dictionaries in isolation from their actual performance in comprehension tasks has less value than it seems (Cumming et al. 1994; Nesi 1998). This is also a reason why in the current study students are requested to do both perception and comprehension tasks.
6. Cumming et al.'s study (1994) demonstrates no significant differences in the effectiveness of definitions (plus illustrative examples) between *COBUILD1* and *LDOCE2* on learners' performance in reception tasks (Cumming et al. 1994). The results seem to answer the question regarding the effectiveness

of two different approaches to controlling a definition vocabulary, but more research is needed in this area.

7. In turn, the study by Nesi (1998) demonstrates that entries from the dictionaries published in the early and mid-1990s (i.e. *OALD5*, *LDOCE3*, *COBUILD2*, *CIDE*, and *COD8*) do not differ with respect to their helpfulness for learners in production or reception (identification) tasks. On the other hand, Nesi and Meara's study (1994), which was performed on earlier dictionaries published in the 1980s, suggests that learners using *OALD4* were significantly more likely to misinterpret their definitions than those who had access to definitions from *LDOCE2* or *COBUILD1*. It would be interesting to explore whether the approach to defining vocabulary is relevant to the comprehension of definitions.
8. The study by McCreary and Dolezal (2006) suggests that in reception tasks native-speaker dictionaries are not more helpful for foreign learners than contextual clues found in a text.

Below are six research questions (RQ) to be considered in the current research. They are provided here together with brief explanations of the motivations for their studies.

RQ1. The current investigation expands on the studies by Nation (2006), Webb and Nation 2008, and Xu (2012) by examining the vocabulary coverage of definitions in a range of different dictionaries, both for learners and for native speakers. The analysis will give us an answer to the question whether dictionaries using a restricted DV offer a lower lexical burden for their readers as compared to those using non-restricted definitions. Differences between the dictionaries in the receptive vocabulary load (RVL) of their definitions (i.e. the knowledge necessary for users to understand the vocabulary of definitions) may suggest that the approach to definition writing (with or without an explicit DV) has an important bearing on the comprehension of definitions. Differences in RVL among native-speaker dictionaries may suggest that the dictionaries pose different levels of difficulty for learners.

RQ2. The core defining vocabularies proposed by Nakane (1998) and Stein (2002b) may no longer pertain to contemporary needs of lexicographic description because they are limited to three dictionaries published over 20 years ago. Since that time, the dictionaries have been revised and new dictionaries have entered the EFL market. It is reasonable, therefore, to determine a core DV on the basis of a larger and updated material (seven lists published between 1987 and 2015). It would be interesting to determine an essential core of a vocabulary that is resistant to change over the time. The question to be answered is how much of the early DVs (1987–1995) survived into the latest dictionaries.

RQ3. In the face of criticism of a restricted DV, in particular by Jansen et al. (1987) and Fox (1989), it is reasonable to expect that contemporary dictionaries implement their defining policies more consistently than in the past. Because over 30 years have passed since *LDOCE1* adopted a DV, one can expect improvements in the practice of controlling definition vocabularies. The current research will investigate to what extent the dictionaries have made improvements in this regard. This also applies to learners' dictionaries that dispense with a restricted DV. It is also worth examining this issue in the context of dictionaries for native speakers. The overview of defining practices in dictionaries published up to the 1980s (in 3.2) shows that a controlled vocabulary is difficult to apply in dictionaries for a wide readership. Finally, it will be interesting to investigate whether digital media take vocabulary control a step further by bringing new solutions to this issue.

RQ4. With respect to user-research, a major focus of the present study is to answer the question to what extent definitions written with a controlled vocabulary are successful in conveying the meaning of a word to the EFL learner. This question will be addressed by a comparison of comprehensibility of invented definitions, written with and without a controlled vocabulary. This step was motivated by the need to eliminate the effect of external factors, such as definition format, definition type, definition length, etc., which might interact with the main variable (the use of a DV). Given previous studies, one may expect that the use of a DV highly significantly contributes to the comprehension score.

RQ5. Another question that needs to be answered is how *authentic* definitions written with and without a DV are perceived and interpreted by learners. Given the fact that previous research in this area is largely confined to learner's dictionaries published up to 1995 (MacFarquhar and Richards 1983; Herbst 1986; Nesi 1998; Cumming et al. 1994; McCreary and Dolezal 1999), research is needed on dictionaries issued in subsequent years. The focus of the current research is on definitions in recent dictionaries intended for learners and for native speakers. Native-speaker dictionaries will be considered in order to examine to what extent their definitions are useful for foreign learners in understanding meaning, and whether such dictionaries are equally challenging for learners. An inspiration for this study was the aforementioned McCreary and Dolezal's conclusion that native-speaker dictionaries offer little help for learners in comprehension tasks (McCreary and Dolezal 1999, 134).

RQ6. A related question is to what extent the variables that are more or less closely related to vocabulary control affect the interpretation of definitions. The variables in question are the approach to defining (with or without explicitly specified DV items), RVL, the percentage of frequent and widely-used words in definitions, definition length, and definition type (i.e. analytical vs. synonym

definitions). It is reasonable to expect that definitions written with a DV are longer than definitions without such restrictions (West 1935, 13; MacFarquhar and Richards 1983; Nesi 2000). As Kirkpatrick explains, “it takes more space to define something in simple terms than it does to define something in more difficult and succinct terms” (1985, 10). As mentioned in 3.1, synonym definitions occupy little space but can be more difficult to understand for learners than analytical definitions (Wiegand 1994; Louw 2000).

The experiments that address questions RQ5 to RQ6 will be conducted on definitions from *LDOCE6* and *COBUILD8*, because they differ in the approach to the selection of vocabulary items for definitions. In addition, definitions from two native-speaker dictionaries will be used: *ODE2* and *Chambers12*. The choice of these dictionaries was prompted by the results of the analysis of the RVL of their definitions, namely the finding that the dictionaries differ with regard to the proportion of high-frequency and wide-range words (see 4.2.1.2).

3.4.2 Research hypotheses

The research questions lead to several hypotheses that are open to experimental study. The hypotheses are shown below together with the symbols for the corresponding research questions (RQ). Hypotheses H1–H4 will be dealt with in Chapter 4, hypothesis H5 in Chapter 4 and 5, and hypotheses H6–H11 in Chapter 6.

A. Hypotheses to be tested in the study of receptive vocabulary load

H1. The RVLs of definitions in the latest editions of dictionaries vary considerably. If this hypothesis proves true for learners’ dictionaries, does the variation depend on the approach to definition writing (with a restricted vs. unrestricted DV)? The hypothesis will be tested separately for:

- a. learners’ dictionaries
- b. native-speaker dictionaries. (RQ1)

H2. Over the recent decades, the receptive vocabulary load of definitions has decreased in:

- a. learners’ dictionaries
- b. native-speaker dictionaries. (RQ1)

H3. Over the recent decades, the percentage of defining words that are potentially most useful and familiar to learners has increased. (RQ1)

B. Hypotheses to be tested in the quantitative and qualitative studies

H4. The DV lists in the early and the recent editions of the learners' dictionaries are to a large extent similar in content. The vast majority of the vocabulary items shared by the early lists has survived into the recent lists. The verification of this hypothesis will help us identify a common vocabulary, which will be referred to as the Core Defining Vocabulary. (RQ2)

H5. The practice of controlling a definition vocabulary has improved over the years. (RQ3)

C. Hypotheses to be tested in the user-oriented study

H6. The contribution of a defining vocabulary to the learner's comprehension of definitions is statistically *highly significant*. (RQ4)

H7. The *LDOCE6* definitions, which are written with a restricted DV, are perceived as being more helpful in explaining the meaning of words than the non-restricted definitions of *COBUILD8*. (RQ5)

H8. The *ODE2* definitions are perceived as being more helpful in explaining the meaning of words than the *Chambers12* definitions. (RQ5)

H9. Definitions in *LDOCE6* and *COBUILD8* are equally comprehensible to foreign learners (RQ5). If this hypothesis is true, does it mean that the approach to a DV has no effect on the comprehension of definitions? (RQ6)

H10. The *ODE2* definitions are more comprehensible for foreign learners than the *Chambers12* definitions. (RQ5)

H11. Comprehension of definitions correlates with the following characteristics of definitions:

- a. RVL
- b. the percentage of frequent and widely-used words in definitions,
- c. definition length,
- d. the percentage of synonym definitions. (RQ6)

The results of the study have implications on the writing of comprehensible definitions. The results will be useful for dictionary editors in making decisions concerning the selection of vocabulary items for definitions and the choice of the approach to definition writing (with restricted or unrestricted DV). By testing the hypotheses we will gain insights into lexicographic practices in this regard and will see how DV policies can be implemented and why they cannot always be

implemented as intended. The verification of the hypotheses will highlight advantages of vocabulary control in definitions and give evidence of its indispensability. The results of the study will be of relevance to lexicographers compiling dictionaries for learners.

CHAPTER 4

Quantitative analyses

In the previous chapters we saw how the pioneering lexicographers set up the models of vocabulary control for learners' dictionaries, and how these models were followed in the early dictionaries. We also saw that in native-speaker dictionaries published up to the early 1980s lexical simplicity of definitions was not as important as their precision and accuracy. This chapter uses quantitative methods to investigate how the policy of vocabulary control developed in the more recent dictionaries. The aim is to verify several hypotheses (H1–H5 presented in Section 3.4.2). One of the fundamental aims is to estimate the lexical burden placed by dictionary definitions on their users, that is, the knowledge that is needed to understand them. A more detailed account of the development of defining vocabularies will be presented in Chapter 5.

4.1 Research design

The approach throughout the experimental part of this book is analytical. The study presented in this section uses several quantitative analyses to examine a selection of dictionaries for EFL learners and for native speakers.

4.1.1 Materials

The learners' dictionaries under study consisted of the British "Big Five", that is, *OALD* (editions 4, 5, and 9), *LDOCE* (2 and 6), *COBUILD* (1 and 8), *CALD* (1 and 4¹), and *MEDAL* (1 and 2); and an American dictionary *MWALED*. All the dictionaries came out between 1987 and 2015. Most of them went through several editions, but the focus of the quantitative study was on the earliest and latest editions published in that period, which was sufficient to test hypotheses H1–H5 and spot general trends in the data. Among the dictionaries under consideration was

1. As is the case with *ODE1*, I am using the abbreviation *CALD1* to refer to the original edition, though in the literature it is normally referred to as *CIDE* (from *Cambridge International Dictionary of English*).

OALD5, though it was not the earliest in the period. I decided to include it because it was the first edition using a restricted DV. In addition, for a broader view of the results, the historical scope was extended by examining several dictionaries published prior to that period, that is, *NMED*, *ISED*, and *LDOCE1*.

As regards native-speaker dictionaries, they were all general-purpose reference works published in the comparable period (between 1984 and 2011): the *Chambers Dictionary* (editions 7 and 12); the *COD* (editions 8 and 12); the *Collegiate Dictionary* (editions 10 and 11); *Oxford Dictionary of English* (*ODE*, editions 1 and 2);² *Longman Dictionary of the English Language* (*LDEL*), and *Oxford English Reference Dictionary* (*OERD*). Most of the dictionaries are well known in the dictionary market; some of them, such as the British *Chambers* and the *COD*, as well as the American *Collegiate* have more than a century-long tradition. All the dictionaries under study were in paper form. For the purposes of this study, definitions from the dictionaries were captured in an electronic format as described in Section 4.1.2.

4.1.2 Data selection

The quantitative study focuses on the examination of definition samples and the DV lists. Definition samples came from a random selection of pages from each dictionary. Each edition (or dictionary) was a source for an independent sample, and no attempt was made to select corresponding definitions from different dictionaries. The latter approach, however, was adopted for the qualitative study (5.2). The number of pages selected from each dictionary ranged approximately from 50 to 80 depending on text density and page size. The hard copies of the pages were converted into text files using an optical scanner. Prior to conversion, definition boundaries were annotated manually, which was necessary for the automatic retrieval of definitions.

Before proceeding further, it may be instructive to explain how the boundaries of definitions were identified. Definitions play a key role in explaining meaning to the user and thus have a central status in dictionary entries. Their central status is rendered visually by using a typeface that is easy to read. Usually definitions appear in normal print to ensure readability, while other entry components are highlighted in a particular way by using italics, boldface, capital letters, or brackets. In this research, the typographical convention served as a guide to identifying the boundaries of definitions. For visual clarity, dictionaries introduce

2. In metalexicography, the title of this dictionary, the *New Oxford Dictionary of English*, is normally abbreviated to *NODE*, but for expository reasons I use *ODE1* in order to disclose the connection with the succeeding *ODE2*.

each definition with a number, letter, or another graphic symbol indicating that the definition corresponds to a lexicographic sense or sub-sense. Such symbols were helpful in this study in identifying definition boundaries but none of them was treated as part of a definition.

The definition has a prominent place at the beginning of the entry, but it is normally preceded and followed by other information types identifying the word being defined. A typical entry (see **valour** and **rub** below) begins with a headword, followed by a variant spelling, a pronunciation form, one or more inflected forms, and labels specifying word class, usage, and grammatical category of the headword. Some of these components are presented in abbreviated form. Another distinct component is an illustrative example, which usually appears in italics.

valour (US **valor**) /'vælə(r)/ *n* [U] (*fml or rhet*) courage, esp in war: *display great valour* (OALD5)

rub /rʌb/, **ruhs**, **rubbing**, **rubbed**. **1** If you **rub** a part of your body, you move your hand back and forward continuously over it while pressing firmly. eg *He groaned and rubbed his eyes ...* (COBUILD1)

Dictionaries usually accord separate status to synonyms, antonyms and other related words by introducing them with labels (e.g., “syn”, “opp”, “compare”) and marking them with a different typeface (e.g., small capitals). In collecting the definition sample, I excluded all the synonyms (and other related items) that were explicitly marked as separate components. This decision was justified by the fact that synonym information does not fulfil the same function as definition (see 3.1). They have distinct but complementary functions, which motivates the exclusion of synonyms from definitions. Yet in preparing the sample, it was not always possible to draw a dividing line between the two types of information, because in practice dictionaries do not always highlight synonyms. For example, in the definition of **valid** “(of arguments, reasons, etc) well based or logical; sound” (OALD5), the word *sound* serves as a synonym information, though it is printed in normal type. Another problem for the researcher is with the word *logical*: should one exclude it because it is a synonym, albeit conjoined with the analytical phrase (“well based”)? The same problem recurs in definitions of most adjectives, which tend to be defined in terms of other adjectives (Atkins and Rundell 2008). Consistency would require that the definition of **term** *v.* “to name, call, or describe” (LDOCE2) be removed entirely, because it is composed exclusively of synonyms. In all these cases, I followed typographic conventions in the dictionaries and included synonyms in the sample, unless they were labelled and set off typographically from definitions.

Definitions sometimes use parenthetical glosses, which either explain a defining word or specify the context of use of the headword (see below). I treated both

types of gloss as integral parts of definitions and included them in the definition sample.

gherkin ... a small type of CUCUMBER (= long, thin, green vegetable) that is often PICKLED (= preserved in VINEGAR) (CALD4)

valid ... (of arguments, reasons, etc) well based or logical; sound (OALD5)

Other intrinsic elements of definitions qualified for inclusion in the definition sample were parenthetical objects of verbs:

cross ... **verb** [with obj.] **1** go or extend across or to the other side of (a path, track, stretch of water, or area) ... (NODE)

On the other hand, excluded from the sample were labels, cross-referenced synonyms and opposites, etymology, usage notes, and other information types that were marked explicitly as separate entry components.

In the initial stage of collecting the definition sample, I did not attempt to analyze definition formats in order to filter out definition parts, such as if-clauses, that had no explanatory value. In other words, full-sentence definitions qualified for inclusion. This approach implied the adoption of the broader concept of definition (3.1), which encompasses not just the *definiens*, that is the explanation of lexical-semantic properties of a lexical unit, but also the *definiendum*, that is the lexical unit which is being defined (Burkhanov 1998). This concept fits the COBUILD model of definition, with the *definiendum* embedded in the contextual part, for example *rub* in the definition “If you **rub** a part of your body ...”.

However, for the analysis of vocabulary load of definitions, it would be misleading to take *definienda* into account, for they are in principle unknown for the user. Therefore, for a valid comparison it was necessary to remove them from the definition sample; otherwise, they would have skewed the results by overstating the receptive vocabulary load. I implemented this precaution not only in COBUILD, which systematically applied full-sentence definitions, but also in the other learners’ dictionaries, which resorted to this strategy less freely than COBUILD.

Having annotated the definition boundaries, I could retrieve definitions and store them in plain files. For the retrieval and subsequent processing of the definitions, I developed a series of R scripts.³ Definitions were checked for errors both manually and automatically. The latter was possible thanks to AntWordProfiler

3. R is an open source programming language (R Development Core Team 2013). It is an object-oriented language allowing a researcher to load, process, and save data in a variety of objects. The simplest form of object is a vector, which may consist of just a character, while the most complex one is a list, which may contain numerous texts, tables, and other data structures (Gries 2009).

and reference lists (see 4.1.3), which highlighted unusual or rare words, allowing for editing the files while loaded into the program. Using the R code, I automatically expanded certain abbreviations which were conventionally used in the dictionaries, such as *esp.*, *usu.*, *sb*, *sth*, *sb/sth*, *in/out*, *specif.*, *orig.*, and *perh.* The slash symbol /, which indicated an alternative in *OALD*, was replaced with a conjunction “or”. Abbreviations that are normally used in general English, such as *e.g.*, *etc.*, *i.e.*, *US*, and *UK*, were retained in their original forms. To ensure a fair comparison, the samples were cleared of all punctuation marks, and then reduced in size to the same number of word tokens: 20,000. This sample size is large enough for a researcher to draw firm conclusions, as it constitutes approximately between 1.4% (*ODE2*) and 11.0% (*NMED*) of dictionary size. For other selected dictionaries, the proportions are as follows: 1.6% in *OERD*, 1.9% in *Chambers7*, 2.1% in *COD8* and *LDOCE6*, 2.3% in *CALD1*, and 6.1% *ISED*. The sample size as measured by the number of definitions ranged from 1115 in *COBUILD1* to 3239 in *Chambers7*.

The comparison of the DV lists was conducted on those learners’ dictionaries that made use of such lists in the period under study (1987–2015). The lists studied came from the following dictionaries: *LDOCE* (editions 2–6), *OALD* (editions 5–9), *CALD* (editions 1 and 4), and *MEDAL* (editions 1 and 2). However, most analyses presented in this chapter were conducted on the earliest and latest editions of these dictionaries. Almost all the dictionaries provide the lists in the back matter. In the case of *CALD*, of which only the first edition came with a DV list, I consulted the publishers, who kindly provided me with the latest list (*CALD4*) as an internal document.⁴ The comparison of the lists in terms of vocabulary load required that all affixes be removed from the lists. Otherwise, they would have been treated as extremely rare items.

When counting the sizes of the lists, I took into consideration all items appearing on the lists. This seems obvious, but we should bear in mind that a researcher comparing word lists is usually confronted with the problem of structural differences among the lists. Some DV lists do not indicate part-of-speech, distinguish between meanings of words, or include complex and multiword items that are used in definitions. Some of the lists under study (e.g., that of *LDOCE2*) have flat structures, while other lists have tired structures, with subheads listed under heads. For example, in *CALD1* there are derivatives *accessible* and *accessibility* nested under the head *access*. In such a case, the count of the size of the list included all heads and nested items.

4. Cambridge Advanced Learner’s Dictionary Defining Vocabulary © 2017. Used by permission of Cambridge University Press.

4.1.3 Research tools and reference lists

The analysis of vocabulary load was conducted with the aid of AntWordProfiler⁵ developed by Laurence Anthony (Anthony 2012). AntWordProfiler is a computer program that generates vocabulary information on a corpus of texts by comparing the corpus against a set of word lists (Anthony 2012). The word lists used in this research were twenty-nine 1,000 word families (henceforth Base Lists) compiled by Nation (2006) on the basis of his research with later modifications (Nation and Webb 2011; Nation 2012).⁶ Each of the Base Lists served as a reference point against which to evaluate the vocabulary load of definitions. This analysis was undertaken to assess how many Base Lists, which consisted of the most useful and familiar words (see further below), were necessary to cover 98% of definition text (see 4.1.4). Word families were made up of derived and inflected forms grouped under a base word irrespective of their lexical classes. Below is a word family for *accept*:

ACCEPT
ACCEPTABILITY
ACCEPTABLE
ACCEPTABLY
ACCEPTANCE
ACCEPTANCES
ACCEPTED
ACCEPTING
ACCEPTOR
ACCEPTORS
ACCEPTS
UNACCEPTABILITY
UNACCEPTABLE
UNACCEPTABLY

In preparing the Base Lists, Nation assigned words to families following several criteria: frequency, regularity, productivity, and predictability of affixes. The criteria were drawn from Bauer and Nation (1993), who used them in arranging affixes into a graded set of seven levels. The levels were designed to serve as the basis for

5. Version 1.4.0.1 for Windows, 2013. Alternatively, this study could have been carried out by the RANGE program developed by Heatley et al. (2002). Both programs take Nation's word-family lists as input for analysis.

6. The lists were downloaded from Paul Nation's website: <https://www.victoria.ac.nz/lals/about/staff/paul-nation#vocab-lists> (last accessed on July 12, 2018). On this website, they appeared as "BNC/COCA word family lists" and came with the RANGE program.

“the staged systematic teaching and learning of these affixes for learners reading English” (Bauer and Nation 1993, 254–255). In arranging affixes into the levels, Bauer and Nation gave priority to regularity of the word formation process for the written form. According to the compilers, the levels represent varying degrees of word transparency. Moving along the levels, transparency of form and meaning decreases. At level 1, which is rather an unlikely scenario, a learner does not recognize that different word forms (e.g., *develop*, *develops*, *developed*, etc.) belong to the same word family, and treats them as separate words. At level 2 it is assumed that a learner recognizes regular inflectional categories of affixes such as plural, 3rd person singular, past tense, and past participle. Level 3 includes the most frequent and regular derivational affixes (e.g., *-able*, *-er*, *-ish*, *-less*, *-ly*, *-ness*, *-th*, and *un-*). When assigning affixes to level 4, the frequency of an affix was more important than its productivity, and the regularity of its orthography was more important than its phonology. This level includes restricted uses of *-al*, *-ation*, *-ess*, *-full*, *-ism*, and *-ist*. Level 5 consists of regular but infrequent affixes (e.g., *-age*, *-al*, *-an*, *-ance*, *-ant*, *-ary*, *-atory*, and *-dom*), whereas level 6 is composed of frequent but irregular affixes (e.g., *-ee*, *-ic*, *-ify*, *-ion*, *-ive*). Classical roots (as in *embolism* and *photography*) and affixes (e.g., *ab-*, *dis-*, *ex-*, *sub-*) belong to level 7.

Word families for Nation’s Base Lists were created with affixes from level 2 to level 6. The goal was to “reflect the likelihood of words being known”, which means that when a learner knows one or two members of a family, “little learning is required for receptive use (comprehension) of other family members” (Nation and Webb 2011, 136–7). Thus, although the compilers included in the first two thousand word lists a number of low-frequency words (e.g., *professional*, *religiosity*, *untranslatable*) as members of word families, the words posed little difficulty in comprehension for the learner (Nation and Webb 2011).

Word families were assigned to the lists on the basis of the range and frequency of occurrence of base words in a corpus (see further below). Range reflects how widely a lexical item is distributed across various sections of a corpus. The parameter is more important than frequency, because a lexical item can be frequent merely in a specific type of text, but rare in other text types. All the word families in Base List 1 to Base List 9 (and partly in Base List 10) had a range of 10, which means that they occurred in each of the ten sections of the corpus (Nation and Webb 2011). The most frequent words were assigned to the early lists. Word families from Base List 10 to Base List 25 were ordered exclusively by frequency, but the words with a range below 3 were excluded.

Schmitt and Schmitt (2014) suggest that the vocabulary between the 4th 1,000 Base List and the 9th 1,000 Base List should be called the mid-frequency vocabulary, and that which goes beyond the latter level should be referred to as low-frequency. They argue that this classification is based on the finding that the

knowledge of the first 9,000 word families is sufficient to read a wide range of authentic texts, ensuring 98% text coverage.

The lists were compiled to be suitable for a variety of purposes: designing courses and vocabulary tests, analyzing the vocabulary of texts, and guiding the creation of simplified texts (Nation and Webb 2011; Nation 2012). A shortcoming of using such lists is that they treat fixed combinations of words (e.g., *take part*) as free combinations, and that they ignore the fact that words have multiple meanings. These problems, however, remain a serious challenge for contemporary computer linguists and scientists working in Natural Language Processing. Because AntWordProfiler is not sensitive to polysemy, the assumption is that defining words are used in their basic meanings.

The Base Lists used in this research consisted of 25 main lists and 4 additional ones (29 in total). The latter were respectively: a list of proper names, a list of marginal words (including letters of the alphabet), a list of transparent compounds, and a list of abbreviations (Nation 2012). These extra lists bore different numbers (from 26 to 29) than Nation's lists (30–33), but they were exactly the same lists as the latter.⁷ For the verification of hypothesis H3, which claims that the percentage of defining words that are potentially most useful and familiar for learners increased over time, it was sufficient to use the first two word lists, with 1,000 headwords each.⁸ This is because the vocabulary they include (the most frequent 1,500–2,000) ensures the greatest return for learning,⁹ that is, “the coverage of text, spoken or written, that knowledge of the words provides” (Nation 1993, 193). It has been shown that after the most frequent 1,500 words, the return for learning vocabulary quickly decreases (Engels 1968; Hwang 1989). As for hypothesis H1, which claims that RVLs of definitions in the latest editions of dictionaries vary considerably, all the 29 lists were used. The first two lists contain the largest word families. The large size of the families in the top lists is due to the fact that higher-frequency stems tend to take a wider range of affixes than lower-frequency stems (Nation 2006, 66). Members of the families are in a transparent relationship with

7. Nation's lists 26–29 were empty and served merely as space for new words. For the correct interpretation of results, I replaced them with Nation's lists 30–33. Thus, the lists 26–29 in the current research correspond to Nation's original lists 30–33.

8. The first and the second list have a total of 2,000 word families, and together with the members of the families the length of the lists reaches 13,227 tokens (6,857 and 6,370 respectively).

9. Gilner (2011) argues that around 2,000 words constitute 70% to 95% of all running words regardless of the source from which the text originates, while Carroll et al. (1971) maintain that the first 1,000 words provide coverage of 74% of written text.

the other members, and the headwords are the most recurring and widely distributed in the language (Nation and Webb 2011).

The composition of the corpus from which the Base Lists were compiled is described in Nation (2012) and Nation and Webb (2011). The first two lists were made from a 10 million token corpus of (mainly) British and American English. The corpus consisted of 10 sections of approximately equal length, drawn from the British National Corpus (BNC), the American National Corpus (AmNC), the Wellington Corpus of Spoken English (New Zealand) (WSC), transcripts of American movies and television programs, and British/New Zealand school journals. The first six sections (6 million tokens) came from spoken English, and the remaining four sections (4 million tokens), from written English. The structure of the corpus used for Base Lists 1 and 2 is as follows (Nation 2012, 1):

1. AmNC spoken face to face, telephone 1
2. AmNC spoken face to face, telephone 2
3. Movies and TV
4. BNC 1
5. BNC 2
6. BNC Plus half of WSC
7. AmNC written fiction, letters 1
8. AmNC written fiction, letters 2
9. School journals
10. BNC fiction

The remaining word lists, from the 3rd to 29th, were compiled on the basis of COCA/BNC rankings in data supplied by Mark Davies (Nation 2012, 2).

The analysis of definition vocabulary with Nation's Base Lists used as a reference point constituted the main part of my quantitative research. Another focus of this research was to compare DV lists with respect to various criteria such as size and similarity of content. To that end, I developed a series of R scripts.

4.1.4 Data analysis

The main goal of the quantitative analysis was to compare the dictionaries with regard to receptive vocabulary load, that is, the amount of lexical knowledge needed by learners to understand definitions (hypotheses H1 and H2). This analysis was inspired by research estimating the effects of text coverage on comprehension. According to Hu and Nation (2000), non-native speakers of English need to know at least 98% of word tokens in a (fiction) text in order to gain "adequate comprehension" of the text, whereas in the study of the academic discourse, Laufer (1989) estimates that 95% can ensure "reasonable reading comprehension" (see

also Carver 1994; Staehr 2009). Assuming that 98% is the ideal coverage that ensures efficient reading of other types of text, including definitions, one can expect that a dictionary user needs to know 98 out of 100 defining words in order to understand definitions. To reach that critical level is a demanding task, though in actual dictionary use, this task is often easier because a learner has access to contextual and grammatical clues that may help the user understand the look-up word (e.g. the original text in which the word occurs, example sentences and grammatical information given in a dictionary). Nevertheless, I adopted the most pessimistic scenario, where these clues are insufficient for the learner to understand the word. Thus, the question of interest was how many word families were needed to reach 98% coverage of definitions. In order to answer this question, each definition sample was loaded into AntWordProfiler and evaluated against the 29 Base Lists. The program generated a range of vocabulary statistics, but our main focus was cumulative percentages of tokens covered by Base Lists. The percentages are shown in Table 2 for learners' dictionaries, and in Table 3 for native-speaker dictionaries.

A similar analysis was carried out to address hypothesis H3, which states that the percentage of familiar and useful words in definitions has increased over the years. This percentage was obtained for each dictionary by running AntWordProfiler with the first two top lists as reference points: Base Lists 1 and 2. These lists lend themselves to this task, as they contain high-frequency and wide-range words as well as numerous derivatives transparently related to base words.

The above analyses, designed to verify hypotheses H1–H3, were performed on two types of data: the definition samples and the DV lists. The results of the former are presented in 4.2.1, whereas those of the latter are in 4.2.2. The decision to analyze different sources of data stems from the fact that an analysis of a DV list can hardly be used as a sole source for drawing conclusions about the vocabulary of definitions. Firstly, not all the dictionaries use such lists; secondly, it is not certain that listed items are really used in definitions (cf. Whitcut 1978); and thirdly, the policy of some dictionaries permits the use of words not listed in their DVs. Therefore, in order to gain more direct insights into the actual practices of controlling vocabularies, the above analyses were conducted on definitions and the lists (the latter when given).

In order to verify hypothesis H4, which claims that the DV lists are similar in content, I conducted a series of quantitative analyses. Using R scripts, I examined the lists from the earliest (1987–1995) and latest editions (2007–2015) in terms of size and the occurrence of various vocabulary items. I computed overlapping and unique items for the early editions and then compared them with those for the recent editions. The lexical overlap is given more attention

Table 1. Input for hierarchical cluster analysis (part of the data).

Vocabulary item	Presence or absence of an item in DV lists							
	<i>OALD5</i>	<i>OALD9</i>	<i>MEDAL1</i>	<i>MEDAL2</i>	<i>LDOCE3</i>	<i>LDOCE6</i>	<i>CALD1</i>	<i>CALD4</i>
a	1	1	1	1	0	1	1	1
a couple	0	1	0	0	0	0	0	0
a few	0	0	0	0	0	0	0	1
a lot	1	0	0	0	0	0	0	1
a.m.	0	1	0	0	0	0	0	0
abandon	1	1	0	0	0	0	0	0

in Section 4.3.3. In order to visualize similarities among the lists, their content was subjected to hierarchical cluster analysis (HCA), a method for dividing a large amount of data into clusters according to similarity. This method has been applied extensively in computational stylometry and forensic linguistics to determine a likely author of a text (Eder et al. 2016; Eder 2017). HCA was performed on a binary set of data arranged in a table, with “0” indicating the absence, and “1”, the presence of a vocabulary item in a list (see Table 1). For example, *a couple* appears only in *OALD9*, and *a few* only in *CALD4*. The first column contains all the items found in all the lists together. Similarity between two given lists was computed on the basis of the absence or presence of a vocabulary item in both lists. The output of the analysis is a tree diagram (dendrogram), showing which lists are similar to each other and at the same time dissimilar to the others (see Section 4.3.2).

The lists were also compared with respect to multiword items and word-building elements (Sections 4.3.5 and 4.3.6). A multiword item is a term that refers to two or more successive words that figure as an entry in a DV list. Thus, elements subsumed under this category include fixed expressions, that is, “lexicalised or coded units” (Moon 1998, 37), collocations, and free combinations of words, as well as strings of words that do not form lexical or syntactic units by themselves (e.g., *about to*). Multiword items were identified automatically on the basis of the presence of spaces or hyphens between words.¹⁰ The

10. The program was unable to identify compounds written as an unbroken stretch of letters (e.g., *cupboard*, *railroad*).

results were grouped under two categories: verbs plus particles/prepositions, and other word combinations (e.g., *card-game*, *about to*, *would like*, *in addition to sth*). As for word-building elements, the lists were compared in terms of the percentage of compound elements and affixes. The count included the elements that *LDOCE* provided in the subordinate list (under the heading “Prefixes and Suffixes”).

4.2 Analysis of receptive vocabulary load and other quantitative comparisons

This section aims to verify hypotheses H1–H4. It is divided into two parts reflecting different types of data under analysis. The first part examines the vocabulary load of definitions, whereas the second one, of the DV lists in learners’ dictionaries.

4.2.1 Receptive vocabulary load of definitions

The analysis presented in this section focuses on differences among dictionaries with respect the lexical burden placed on their users. The advantage of this analysis is that it focuses on the *actual* vocabulary to which the user is exposed while reading definitions. This is important especially for the analysis of those dictionaries that use restricted DVs and occasionally use words outside their lists.

4.2.1.1 Coverage of definitions by Base Lists: learners’ dictionaries

Table 2 shows the results of the analysis of vocabulary load in learners’ dictionaries. For comparative purposes, the table also includes data for dictionaries published prior to 1984, such as *NMED*, *ISED*, and *LDOCE1*. Each row in the table displays a cumulative percentage of defining words (word tokens) covered by respective Base Lists. Recall that each Base List contains 1,000 word families, and a word family consists of a headword in the form of a base word and all its inflected and derived forms that are easy to understand for a learner who knows the base word (see 4.1.3). Thus, for instance, the first Base List covers 86.74% word tokens in the *NMED* definitions, the first and the second Base Lists cover in total 94.82% tokens, and the first three Base Lists bring the total coverage to 96.74%. The first five Base Lists, which are highlighted in grey, contribute to 98% text coverage in this dictionary.

Investigating the latest editions of the learners’ dictionaries, one notices that the vocabulary needed to cover 98% of the examined definition texts ranges from 3,000 word families in *LDOCE6* and *CALD4*, to 4,000 in *OALD9*, *COBUILD8*, *MEDAL2*, and *MWALED*. In other words, the dictionaries have a light receptive

vocabulary load, as only 3 to 4 Base Lists are needed to reach 98% coverage of defining words. Importantly, this coverage is reached within 3 lists only by those dictionaries (*LDOCE6* and *CALD4*) that use a restricted DV. This is an interesting observation, which suggests the superiority of this approach to definition writing. However, note that, as far as the latest editions are concerned, this approach does not directly lead to a lower vocabulary load, as the other dictionaries which use such restrictions (i.e. *OALD9* and *MEDAL2*) need one list more to reach that coverage (Base Lists 1 to 4). Nevertheless, all the dictionaries display comparable levels of receptive vocabulary load. This argues against hypothesis H1, which claimed that the RVLs of definitions in the latest editions of dictionaries vary considerably.

Analyzing changes in vocabulary load between 1987 and 2015, one notices that this parameter decreased in most dictionaries studied, that is *OALD*, *LDOCE*, and *COBUILD*. The sharpest drop was noted in *OALD*, from 10,000 word families in *OALD4* to 5,000 in *OALD5*. It must be noted that this decline was arguably due to the fact that the latter edition adopted a limited DV for the first time in the history of this dictionary. This was a radical change in the editorial policy, which was likely to contribute to definitions being easier for the learner to understand.

The tendency to reduce the vocabulary load in *OALD* is not by any means a recent phenomenon. If one takes *ISED* (1942) as a reference point, the decrease is even more noticeable, as this dictionary has an exceptionally high vocabulary load. As Table 2 demonstrates, *ISED* needs 14 Base Lists to cover 98% of the definition vocabulary. The *ISED*'s heavy lexical burden contrasts markedly with the relatively light load of *NMED*, which needs only 5 Base Lists to reach 98% coverage. Although the dictionaries were designed for different target audiences (advanced in *ISED* and intermediate in *NMED*), one cannot exclude the possibility that the approach to definition writing (unrestricted vs. restricted DV) may have an effect on vocabulary load. This is clear when one compares the vocabulary loads of two advanced learners' dictionaries published in the 1980s: *OALD4* and *LDOCE2* (10 Base Lists vs. 6 Base Lists, respectively).

With regard to words potentially familiar to learners, which are covered by Base Lists 1 and 2, their proportions increased over the years in most dictionaries. As Table 2 shows, the largest increase was observed in *OALD*, from 86.76% in *OALD4* to 93.89% in *OALD9*. This was a considerable change (by more than 7%), which arguably resulted from the aforementioned new approach to vocabulary control. In two other dictionaries, increases are less significant, from 92.74% in *LDOCE2* to 95.57% in *LDOCE6*; and from 93.62% in *COBUILD1* to 95.39% in *COBUILD8*. There are practically no changes in *CALD* or *MEDAL*.

Table 2. Cumulative percentage coverage of definitions in learners' dictionaries by Base Lists

Base List (1,000 families each)	<i>NMED</i> (1935)	<i>ISED</i> (1942)	<i>OALD4</i> (1989)	<i>OALD5</i> (1995)	<i>OALD9</i> (2015)	<i>LDOCE1</i> (1978)	<i>LDOCE2</i> (1987)	<i>LDOCE6</i> (2014)
1	86.74	76.69	76.00	79.91	84.88	82.82	82.00	87.16
2	94.82	87.31	86.76	90.47	93.89	92.69	92.74	95.57
3	96.74	92.78	93.31	96.32	97.41	96.00	96.30	98.25
4	97.53	94.82	95.29	97.52	98.08	97.02	97.15	98.72
5	98.11	96.02	96.42	98.11	98.50	97.89	97.95	98.98
6	98.30	96.70	97.03	98.50	98.71	98.20	98.27	99.13
7	98.43	97.06	97.45	98.68	98.90	98.48	98.55	99.23
8	98.48	97.35	97.73	98.78	98.98	98.68	98.68	99.29
9	98.51	97.58	97.94	98.86	99.04	98.77	98.80	99.36
10	98.51	97.74	98.13	98.91	99.09	98.84	98.89	99.38
11	98.52	97.87	98.26	98.98	99.14	98.91	98.96	99.42
12	98.52	97.91	98.35	99.00	99.17	98.95	98.99	99.44
13	98.54	97.96	98.38	99.03	99.19	99.00	99.02	99.45
14	98.54	98.02	98.43	99.03	99.22	99.02	99.05	99.46
15	98.55	98.03	98.50	99.03	99.24	99.03	99.06	99.47
16	98.55	98.04	98.51	99.03	99.26	99.05	99.07	99.47
17	98.55	98.06	98.52	99.03	99.28	99.08	99.07	99.49
18	98.55	98.07	98.53	99.03	99.30	99.09	99.07	99.49
19	98.55	98.08	98.54	99.03	99.32	99.10	99.08	99.49
20	98.55	98.09	98.54	99.03	99.33	99.11	99.08	99.50
21	98.55	98.09	98.54	99.03	99.33	99.11	99.08	99.50
22	98.55	98.09	98.54	99.03	99.33	99.11	99.08	99.50
23	98.55	98.09	98.54	99.03	99.33	99.11	99.08	99.50
24	98.55	98.09	98.54	99.03	99.34	99.11	99.08	99.50
25	98.55	98.10	98.56	99.03	99.34	99.11	99.08	99.50
26	99.11	99.06	99.17	99.31	99.81	99.53	99.46	99.79
27	99.19	99.31	99.28	99.37	99.87	99.58	99.52	99.80
28	99.32	99.57	99.59	99.45	99.97	99.78	99.74	99.91
29	99.90	99.85	99.92	99.94	99.97	99.79	99.83	99.94
Not in the lists [%]	0.08	0.12	0.07	0.01	0.03	0.15	0.13	0.01

Table 2. (Continued)

Base List (1,000)	COBUILD1 (1987)	COBUILD8 (2014)	CALD1 (1995)	CALD4 (2013)	MEDAL1 (2002)	MEDAL2 (2007)	MWALED (2008)
1	85.84	87.20	86.40	86.10	87.14	86.02	85.07
2	93.62	95.39	95.56	95.84	95.49	95.04	94.49
3	97.00	97.83	98.12	98.20	97.73	97.76	97.54
4	97.99	98.42	98.80	98.77	98.38	98.36	98.32
5	98.49	98.91	99.13	99.01	98.64	98.77	98.69
6	98.68	99.06	99.31	99.16	98.80	98.95	98.87
7	98.92	99.20	99.39	99.23	98.93	99.09	98.98
8	99.02	99.27	99.47	99.30	99.00	99.14	99.13
9	99.10	99.31	99.50	99.34	99.06	99.20	99.21
10	99.16	99.35	99.52	99.37	99.10	99.24	99.25
11	99.21	99.38	99.55	99.39	99.11	99.29	99.26
12	99.25	99.41	99.56	99.39	99.13	99.31	99.27
13	99.27	99.41	99.57	99.42	99.17	99.34	99.29
14	99.29	99.42	99.58	99.42	99.20	99.37	99.32
15	99.30	99.42	99.58	99.42	99.20	99.37	99.34
16	99.31	99.43	99.58	99.42	99.20	99.37	99.35
17	99.33	99.43	99.59	99.42	99.20	99.38	99.36
18	99.33	99.43	99.59	99.42	99.21	99.38	99.36
19	99.33	99.43	99.59	99.42	99.21	99.38	99.36
20	99.33	99.44	99.59	99.42	99.21	99.38	99.37
21	99.33	99.44	99.59	99.42	99.21	99.38	99.37
22	99.33	99.44	99.59	99.42	99.21	99.38	99.37
23	99.33	99.44	99.59	99.42	99.21	99.38	99.37
24	99.33	99.44	99.59	99.42	99.21	99.38	99.37
25	99.33	99.44	99.59	99.42	99.21	99.38	99.37
26	99.84	99.75	99.82	99.75	99.65	99.70	99.71
27	99.88	99.78	99.85	99.77	99.71	99.72	99.74
28	99.92	99.89	99.93	99.92	99.86	99.82	99.85
29	99.93	99.92	99.94	99.95	99.93	99.91	99.87
Not in the lists [%]	0.01	0.03	0.01	0.01	0.02	0.04	0.07

4.2.1.2 Coverage of definitions by Base Lists: native-speaker dictionaries

As could be expected, the definitions of the native-speaker dictionaries demonstrate much heavier vocabulary loads than those of learners' dictionaries (Table 3). However, there is little variation among the native-speaker dictionaries with respect to the number of word families needed to reach 98% coverage: from 26,000 to 28,000. Among the latest editions, only *COD12* and *ODE2* reach this threshold with 26,000 word families.

From a diachronic point of view, changes in vocabulary load are relatively minor. The Chambers definitions have become lexically richer by 2,000 word families, whereas there is no change in *COD* or *ODE*. The *Collegiate* is the only dictionary whose definitions have become lexically lighter over the years, but the change (by 1,000 word families) is rather negligible.

Understandably, all the dictionaries under study use a certain proportion of low-frequency words. If one considers low-frequency words from the lists 10 to 25 (see 4.1.3), their proportions range from 1% in *NODE* to 2.3% in *Chambers7*. In addition, there are rare words that do not appear on any Base List (from 0.77% in *COD8* to 1.61% in *Collegiate10*), which may present a stumbling block for less competent users.

If we examine the proportion of high-frequency and wide-range words covered by the first two 1,000 word families, a greater variation emerges among the dictionaries. The values range from 76.60% in *Collegiate10* to 82.55% in *NODE*. Both editions of *COD* and of *ODE* appear to have a relatively greater proportion of such words (above 79%), as compared to the other dictionaries (less than 79%). This means that their users are exposed to a greater number of common words while reading definitions, which may have a positive effect on their comprehension. However, there is no convincing evidence that recent editions use a greater number of high-frequency and wide-range words than earlier editions do, as the percentage of such words increased by 1.2% in *COD* and decreased by 3.5% in *ODE*.

4.2.2 Receptive vocabulary load of DV lists

In general, lexicographers compile DVs by selecting those lexical units that are most useful for learners in reading definitions and for definers in explaining meanings. To what extent this policy produces the same results is an interesting question that will be addressed in this section. This issue will be dealt with by a comparison of the DV lists with regard to size and lexical content. A related issue is the composition of the vocabulary that is shared by all the lists. What are defining words on which editors agree, and how many of them are there in the lists?

These questions will be answered in Sections 4.3.3 and 4.3.4. First, however, we need to estimate the receptive vocabulary load of the lists. This examination will be conducted in the same way as we examined definition samples in Section 4.2.1.

4.2.2.1 Coverage of DV lists by Base Lists

Table 4 presents percentage coverages of each DV list by Nation's Base Lists. Considering the latest editions, one finds that the majority of the dictionaries (i.e. *OALD9*, *LDOCE6*, and *CALD4*) reach 98% coverage with the first 5 Base Lists, whereas *MEDAL2* needs 7 Base Lists. In other words, the recent DV lists have identical or comparable levels of difficulty for learners. Interestingly, the early editions display the same or comparable vocabulary loads, with the exception of *CALD1*'s list, which needs 12 Base Lists to reach 98%. This list appears to be more difficult for learners to master than the other lists.

However, when we look at the proportion of the most frequent and useful words, which are covered by Base Lists 1 and 2, we will see that it has increased in the majority of the dictionaries. The greatest increase is found in *OALD*: 8.21%, while the respective values for *LDOCE* and *CALD* are 4.18% and 6.51%. *MEDAL* does not show a significant change in this regard. Among the recent editions, the highest proportion is in *LDOCE6* (85.66%), and the lowest in *OALD9* (80.51%). These figures are a sign of improvement, as users of recent editions have less chance of encountering unknown words in definitions.

As Table 4 shows, the sharpest drop in vocabulary load is in the *CALD* lists, from 12,000 to 5,000 word families. A close inspection of the lists indicates that there are at least two reasons behind the decline. The first one is the increase (by more than 6%) in the percentage of the most frequent words which are covered by Base Lists 1 and 2, and the second is the removal of a number of less frequent but semantically transparent derivatives covered by Base Lists 7–12, for example *habitual*, *inseparable*, *inventive*, *misfortune*, *repetitive*, *electrify*, *trustworthy*, *cylindrical*, *reflector*, *sociable*, *weaponry*, *disobey*, *acidified*, *indivisible*, and *springy*. As a result, the learning burden of the *CALD4* list is lighter than that of *CALD1*. We will have a closer look at the *CALD* policy towards derivatives in Sections 4.3 and 5.2.4.

It should be noted, however, that the above comparison is tentative because some dictionaries allow for the use of words that are not on the lists, while others follow a more restrictive policy. For example, *LDOCE* allows lexicographers to use derivatives formed by means of affixes from the list, whereas *MEDAL* permits only those derivatives that are on the list (see 5.2). Such discrepancies will be discussed in more detail in Section 4.3.

Table 3. Cumulative percentage coverage of definitions in native-speaker dictionaries by Base Lists

Base List (1,000)	Chambers7 (1988)	Chambers12 (2011)	COD8 (1990)	COD12 (2011)	Collegiate10 (1993)	Collegiate11 (2004)	LDEL (1984)	NODE/ODE1 (1998)	ODE2 (2003)	OERD (1996)
1	67.59	66.98	67.85	69.05	65.56	66.19	66.47	71.35	67.63	66.10
2	78.33	77.67	79.44	80.63	76.60	77.62	78.38	82.55	79.04	77.09
3	84.72	85.40	87.77	87.94	83.64	85.01	86.26	88.77	86.27	84.79
4	87.97	88.33	90.71	90.69	86.82	87.80	89.29	91.26	89.24	87.63
5	90.05	90.24	92.62	92.50	88.65	89.70	91.03	92.81	90.93	89.29
6	91.41	91.74	93.90	93.85	90.14	91.18	92.25	94.08	92.09	90.39
7	92.42	92.68	94.78	94.58	91.31	92.05	93.08	94.76	92.75	91.20
8	93.45	93.53	95.49	95.27	92.22	92.83	93.93	95.37	93.29	91.96
9	94.12	94.11	95.98	95.70	92.81	93.30	94.41	95.70	93.73	92.32
10	94.56	94.60	96.41	96.04	93.26	93.65	94.76	96.05	94.08	92.69
11	95.06	94.92	96.65	96.29	93.60	94.06	95.11	96.22	94.41	92.99
12	95.42	95.26	96.99	96.48	93.94	94.27	95.42	96.35	94.68	93.25
13	95.79	95.49	97.14	96.61	94.18	94.50	95.68	96.49	94.82	93.45
14	96.03	95.65	97.24	96.76	94.30	94.72	95.80	96.64	94.95	93.55
15	96.16	95.82	97.31	96.83	94.39	94.83	95.94	96.72	95.08	93.62

Base List (1,000)	Chambers7 (1988)	Chambers12 (2011)	COD8 (1990)	COD12 (2011)	Collegiate10 (1993)	Collegiate11 (2004)	LDEL (1984)	NODE/ODE1 (1998)	ODE2 (2003)	OERD (1996)
16	96.36	95.96	97.37	96.95	94.54	94.97	96.06	96.77	95.19	93.74
17	96.44	96.06	97.47	97.03	94.65	95.16	96.14	96.82	95.31	93.86
18	96.56	96.09	97.50	97.11	94.71	95.29	96.22	96.87	95.35	93.91
19	96.63	96.14	97.53	97.18	94.75	95.40	96.30	96.92	95.40	93.96
20	96.67	96.22	97.56	97.21	94.87	95.49	96.33	96.95	95.44	93.99
21	96.74	96.26	97.58	97.26	94.95	95.55	96.34	97.00	95.48	94.00
22	96.80	96.32	97.61	97.29	94.99	95.60	96.36	97.01	95.51	94.02
23	96.83	96.35	97.62	97.30	95.07	95.73	96.37	97.04	95.54	94.05
24	96.01	96.36	97.66	97.31	95.10	95.74	96.39	97.05	95.56	94.09
25	96.89	96.39	97.71	97.32	95.18	95.78	96.42	97.06	95.60	94.13
26	98.01	97.68	98.65	98.38	96.21	96.86	97.47	98.28	98.49	97.91
27	98.07	97.92	98.75	98.44	97.89	98.33	97.80	98.34	98.53	98.02
28	98.48	98.31	99.10	98.99	98.32	98.67	98.18	98.73	99.04	98.50
29	98.62	98.54	99.20	99.10	98.36	98.77	98.95	98.77	99.22	98.67
Not in the lists [%]	1.34	1.44	0.77	0.87	1.61	1.20	1.02	1.20	0.86	1.36

Table 4. Cumulative percentage coverage of DV lists by Base Lists (affixes excluded).

Base List (1,000)	OALD5 (1995)	OALD9 (2015)	LDOCE2 (1987)	LDOCE6 (2014)	CALD1 (1995)	CALD4 (2013)	MEDAL1 (2002)	MEDAL2 (2007)
1	41.87	53.33	50.86	53.98	47.24	54.09	52.68	52.55
2	72.30	80.51	81.48	85.66	77.37	83.88	83.25	83.01
3	89.76	94.68	90.22	94.34	87.87	93.01	93.29	93.15
4	96.31	97.55	95.47	97.92	92.94	96.68	96.58	96.52
5	98.39	98.43	97.51	98.82	95.04	98.02	97.57	97.51
6	98.84	98.80	98.23	99.10	96.21	98.38	97.93	97.87
7	99.12	98.94	98.64	99.29	97.03	98.70	98.13	98.11
8	99.32	99.03	98.82	99.34	97.43	98.82	98.25	98.19
9	99.38	99.03	98.91	99.34	97.72	98.82	98.25	98.19
10	99.41	99.03	98.96	99.39	97.85	98.86	98.33	98.27
11	99.44	99.08	99.05	99.44	97.98	98.86	98.37	98.31
12	99.44	99.08	99.05	99.44	98.09	98.90	98.41	98.35
13	99.44	99.10	99.05	99.44	98.09	98.90	98.41	98.35
14	99.44	99.10	99.10	99.49	98.12	98.94	98.45	98.39
15	99.44	99.10	99.10	99.49	98.12	98.94	98.45	98.39
16	99.44	99.10	99.10	99.49	98.17	98.94	98.45	98.39
17	99.44	99.10	99.10	99.49	98.17	98.94	98.45	98.39
18	99.44	99.10	99.10	99.49	98.17	98.94	98.45	98.39
19	99.44	99.10	99.10	99.49	98.17	98.94	98.45	98.39
20	99.44	99.10	99.10	99.49	98.17	98.94	98.45	98.39
21	99.44	99.10	99.10	99.49	98.17	98.94	98.45	98.39
22	99.44	99.10	99.10	99.49	98.17	98.94	98.45	98.39
23	99.44	99.10	99.10	99.49	98.17	98.94	98.45	98.39
24	99.44	99.10	99.10	99.49	98.17	98.94	98.45	98.39
25	99.44	99.10	99.10	99.49	98.17	98.94	98.45	98.39
26	99.47	99.10	99.15	99.49	98.17	98.94	99.05	99.02
27	99.47	99.24	99.20	99.54	98.22	98.98	99.09	99.06
28	99.98	99.96	100	100	98.78	99.81	99.96	99.93
29	99.98	99.98	100	100	98.78	99.85	100	99.97
Not in the lists	0.02	0.02	0	0	1.22*	0.15	0	0.03

*This count includes two misprints: *severly* and *worsh* (instead of *worship*) that appear on the list.

4.3 Similarity of content

This section is designed to verify hypothesis H4, which claims that the DV lists in the early and the recent editions are largely similar in content. To test the hypothesis, we perform a series of analyses that compare the lists in terms of measurable parameters, such as size, the number of overlapping and unique items, and the proportion of multiword items and word-building elements. The results are presented in Tables 5–13 and Figures 2–3. The common vocabulary is analyzed in more detail in Section 4.3.4.

4.3.1 Size of the lists

Figure 2 shows the sizes of the DV lists in the earliest and latest editions of each dictionary. The majority of the latest lists are relatively short, ranging from 2,128 items in *LDOCE6* to 2,472 in *CALD4*. The *OALD9* list, which has 3,876 items, is almost twice as long as that of *LDOCE6*.

The changing sizes of the lists suggest that the vocabularies have undergone significant revision. The *CALD* list in its early edition relied on a very long list (3,750 items). Over the years, the list has been reduced radically by nearly 1,300 items, to the size of 2,472 items. As for *OALD*, the list has increased in size by 389 items compared with the original *OALD5* list. The latter was reduced (from 3,487) to 2,946 in *OALD6*, and then extended to 3,785 in *OALD7*, where it was called “the Oxford 3000”. The list was expanded further in *OALD9*, reaching 3,876 items. In turn, the *MEDAL* list has changed the least (by 8 items), but this is hardly surprising, given the short interval between the publication dates of the *MEDAL* editions (i.e. 2002–2007).

Comparing the successive lists of *LDOCE*, one notices that they fluctuated in size from 2,249 in *LDOCE2* to 2,092 in *LDOCE3*. The greatest revisions are observed in the second and the third edition (cf. Xu 2012).¹¹ While, *LDOCE2* added 226 items and removed 199 from the original list (2227 items), *LDOCE3* added 178 items and removed 331 from the previous list, and the respective figures for *LDOCE4* are 44 and 26. In *LDOCE5*, only 2 items were removed, while in *LDOCE6*, 27 were added and 9 were removed. Thus, the revision of the *LDOCE* list was not as great as that of *OALD*’s or *CALD*’s.

11. The respective figures may be different from Xu’s. One of the reasons is that the current study takes into consideration affixes from the subordinate list (in *LDOCE*).

The above numbers, however, do not always tally with what editors claim about vocabulary sizes. For example, definitions in *CALD1* are claimed to have been “written using a list of under 2,000 words” (*CALD1*, 1702), a figure which is perhaps based on the calculation of headwords rather than nested derivatives and multiword expressions. A more detailed analysis of the content of DVs (see further below) will show that the dictionaries do not follow a uniform policy towards the categories selected for inclusion. For example, the *OALD9* list includes part-of-speech homographs such as *can v.* and *can n.* as separate vocabulary items, as well as free combinations of words such as *think of as*, *make progress*, and *turn around*, which are excluded from the other dictionary lists. In turn, *LDOCE6* enters affixes, such as *-able*, *-al*, *-ance*, *-ation*, *-ed*, *-ence*, *-er*, *-ful*, *-ic*, *-ical*, *-ing*, *-ion*, etc., which are missing from the other dictionary lists. Thus, differences between the DVs are considerably greater than they may seem from a comparison of their sizes. It appears that the length of a DV is not a reliable measure of the actual content of definitions and needs to be complemented with other criteria.

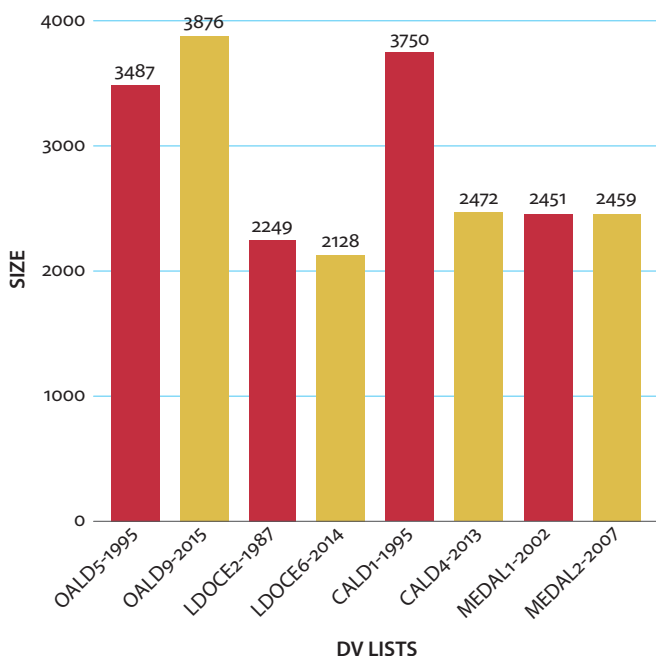


Figure 2. The size of the DV lists in the earliest and the latest editions published in 1987–2015.¹²

12. For comparative purposes, the count of the DV size pertains to all word-forms, not lexemes. It covers word forms provided on the same line of the list. Thus, items given in parentheses or after a comma are counted as separate; for example, “*worthy (of)*”, “*actor, actress*” (in *LDOCE2*) are treated as *worthy*, *worthy of*, *actor*, *actress*, respectively. The count also covers

4.3.2 HCA: Similarities between the lists

Hierarchical Cluster Analysis was run on the data shown in part in Table 1. The results of the analysis take the form of a dendrogram (Figure 3), which is a rough illustration of similarities between the lists. In the dendrogram, the DVs with the most similar lexical content merge into pairs. The level at which they merge is meaningful: the lower the level, the stronger the similarity between the lists.

At the bottom of the diagram, one finds two editions of *MEDAL*, which means that they are most alike. They are joined at a higher level by *CALD4* to form a larger cluster, showing that the *CALD4* list is more similar to that of *MEDAL* than to the other lists. Note that the early list, *CALD1*, is entirely different, as it is more similar to *OALD*'s lists than to *CALD4*'s list. Another pair is formed by two editions of *LDOCE* at a somewhat higher level, the level which apparently reflects a long period that has passed since *LDOCE2* (1987–2014). The division of the dendrogram into two main branches reflects the major dissimilarities between the dictionaries. The right-hand branch is essentially made up of two editions of *OALD*, which indicates their similarity.

The results of the cluster analysis can be summarized as follows: the latest lists of *LDOCE*, *MEDAL*, and *OALD* are a continuation of the early lists, whereas *CALD4* shows a radical departure from the early tradition. The vocabularies of *LDOCE*, *MEDAL*, and *OALD* have been relatively stable, in contrast to *CALD*'s.

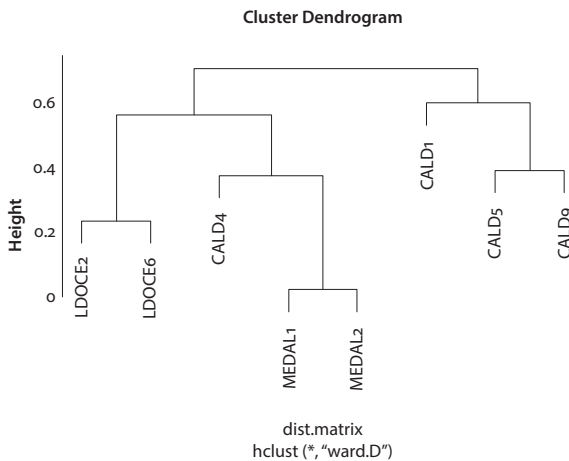


Figure 3. Hierarchical cluster analysis: Visualization of similarity of DV lists.¹³

affixes which are provided in the subordinate list of *LDOCE2*. Phrasal verbs and other word combinations are treated as whole items (e.g., *look for*). Homographs (*wind* v. and *wind* n.) are also counted as separate items.

13. The input data (see Table 1) was processed with *dist.matrix* function in R, with the distance method set to “binary” to match the binary type of the data. The dendrogram was

4.3.3 Overlapping and unique items

Table 5 displays the proportions of vocabulary items that are common to the DV lists in the early editions, published between 1987 and 1995, and Table 6 shows items common to the lists in the latest editions, published in 2007–2015. The tables also provide proportions of unique items, which are found only in one out of the early lists (Table 5), and one out of the recent lists (Table 6). The early editions share 1739 vocabulary items, and the recent ones, exactly the same number. A further comparison of these two groups of items shows that they are predominantly the same, which means that they have survived over the years. This vocabulary will be described in more detail in Section 4.3.4.

Taking into consideration the early dictionaries, one finds that the overlapping items account for a relatively large proportion of each dictionary list, from 46% in *CALD1* to 77% in *LDOCE2* (Table 5). The respective figures for the latest editions are 45% in *OALD9* and 82% in *LDOCE6* (Table 6). It is worth noting that, among the recent editions, the largest overlap is found in *LDOCE6* (82%), *MEDAL2* (71%), and *CALD4* (70%), which makes these lists alike. This finding coincides with the results of the cluster analysis (cf. Figure 3). Another finding is that there was a good deal of editorial experimentation with the early lists, especially those of *CALD1* and *OALD5*, because they contain a large number of unique items (36% and 25%, respectively). All the lists under study have been revised, but the *CALD* list has become very similar to that of *LDOCE* and *MEDAL*, whereas the *OALD* list, with 33% of unique items, have retained its distinctive character.

Further analysis of the development of the lists over the years (Tables 7–10) shows that the *MEDAL* list has changed least of all, which follows from the fact that it has retained the largest proportion of the original vocabulary (99%). In *OALD* and *LDOCE*, this proportion is relatively high (79% and 80%, respectively), whereas in *CALD*, it is significantly lower (60%). The latter finding results chiefly from the removal of numerous derivatives from the original list (see 5.2.4).

Table 5. Overlapping and unique vocabulary items in the early DVs.

Dictionary	Unique items	Overlapping items	
<i>OALD5</i> (1995)	25% (861)	50%	
<i>LDOCE2</i> (1987)	7% (156)	77%	1739
<i>CALD1</i> (1995)	36% (1368)	46%	

produced with the function *hclust* using the linkage method “ward.D”. This method has been shown to perform well in corpus linguistics (Eder 2017).

Table 6. Overlapping and unique vocabulary items in the recent DVs.¹⁴

Dictionary	Unique items	Overlapping items	
<i>OALD9</i> (2015)	33% (1289)	45%	
<i>LDOCE6</i> (2014)	3% (67)	82%	1739
<i>CALD4</i> (2013)	8% (194)	70%	
<i>MEDAL2</i> (2007)	5% (123)	71%	

Table 7. Overlapping and unique vocabulary items in the *OALD* DVs.

Dictionary	Unique items	Overlapping items	
<i>OALD5</i> (1995)	20% (708)	79%	2766
<i>OALD9</i> (2015)	28% (1071)	71%	

Table 8. Overlapping and unique vocabulary items in the *LDOCE* DVs.

Dictionary	Unique items	Overlapping items	
<i>LDOCE2</i> (1987)	20% (436)	80%	1778
<i>LDOCE6</i> (2014)	16% (348)	84%	

Table 9. Overlapping and unique vocabulary items in the *CALD* DVs

Dictionary	Unique items	Overlapping items	
<i>CALD1</i> (1995)	39% (1456)	60%	2262
<i>CALD4</i> (2013)	7.8% (193)	91%	

Table 10. Overlapping and unique vocabulary items in the *MEDAL* DVs

Dictionary	Unique items	Overlapping items	
<i>MEDAL1</i> (2002)	1% (25)	99%	2424
<i>MEDAL2</i> (2007)	1% (33)	99%	

14. The data for the *MEDAL1* list (2002) is not given in the table, because the overlap (71%) is identical to that of *MEDAL2* (2007).

In Section 4.3.4, we will see that the dictionaries agree on the selection of words that are most frequent in the language, widely distributed across various lexical fields, useful in defining other words, and easy for learners to understand. However, the selection is not always straightforward.

Some vocabulary items are more questionable than others. A comparison of the DV lists from the latest editions shows that numerous words appear in one or more lists but not in the others. Table 11 provides examples of vocabulary items that occur only in one of the lists.

Table 11. Examples of vocabulary items unique to a particular DV list.

<i>LDOCE6</i>	<i>-able, -al, -ance, as opposed to, carriage, cattle, comb, cruelty, duck, favourable, fog, foreigner, in spite of, kneel, lamb, mist, nasty, rabbit, rat, slippery, teenage, temper, tribe</i>
<i>OALD9</i>	<i>abandon, account for, because of, beef, bring down, burst into, clerk, come in, conference, hearing, household, incident, item, jealous, jeans, joy, mild, photo, shave, umbrella</i>
<i>CALD4</i>	<i>candle, cave, cone, hunger, idiom, mammal, organism, oval, pastry, sausage, ski, reptile, slang, tissue, syllable, tissue, usefulness, whitish</i>
<i>MEDAL2</i>	<i>Anglican, Buddhism, Catholic, Christianity, communism, democracy, download, elephant, gene, kidney, learner, missile, Muslim, orchestra, Protestant</i>

Some of the above items are easily replaceable by other words. A few examples will illustrate this point. The *MEDAL2* definition of **hostel** uses “at low prices” rather than just “cheaply” apparently because the latter is not on the list. In turn, *CALD4* defines **hostel** in terms of “cheaply”, a word from the list:

- hostel** a building where people living away from home can stay and get meals at low prices (MEDAL2)
- hostel** a large house where people can stay free or cheaply (CALD4)

MEDAL2 dispenses with *brightly*, because it is not on the list, but expresses the meaning of the word using a paraphrase *full of light or colour*:

- ablaze** bright and full of light or colour (MEDAL2)
- ablaze** brightly lit or brightly coloured (CALD4)

All the dictionaries but *CALD4* have *husband* and *girlfriend* on their lists, and while *LDOCE6* defines **fidelity** in terms of these words, *CALD4* does so in terms of *partner*:

- fidelity** “when you are loyal to your husband, girlfriend etc., ...” (LDOCE6)
- fidelity** “honest or lasting support, or loyalty, especially to a sexual partner” (CALD4)

However, it is not always the case that a non-DV item can easily be replaced by or paraphrased with the words from the list. For example, names of religions and their followers like *Buddhism*, *Catholic*, *Christianity*, *Muslim*, and *Protestant* appear in definitions of culture-specific terms¹⁵ in all the dictionaries under study, but only *MEDAL2* explicitly places them in its list. To take another example, all the dictionaries use *jeans* in definitions of **denim** and **Levis**, but only *OALD9* enters it in the list. Other examples are *jealous* and *umbrella*. The former is not in the *LDOCE6* list, though it is used in the definition of **bitter**; and the latter does not appear on the *CALD4* list, but is used in the definition of **sunshade**:

bitter “feeling angry, jealous, and upset ...” (*LDOCE6*).

sunshade “an object similar to an umbrella that you carry to protect yourself from light from the sun” (*CALD4*).

Note that in these definitions the non-DV words figure in normal type, with no indication of their outsider status (this topic will be discussed further in Section 5.2). Arguably, certain low-frequency but precise words are difficult to avoid in some definitions.

Finally, the dictionaries do not agree on the selection of basic-level grammatical categories. Words like *abbreviation*, *adjective*, *adverb*, *noun*, and *verb* figure on all the latest lists but *OALD9*'s. Yet, interestingly, *OALD9* does use them in the definitions of **comparative**, **intensifier**, **abstractnoun**, and **clause**.¹⁶ Thus, the words are necessary in definitions of more specific grammatical concepts.

4.3.4 The Core Defining Vocabulary

As mentioned in 4.3.3, the early lists share 1739 items, and the recent lists share the same number. The overlap between these two groups is significantly large: 1,520 items (87%) – these are the items that have survived into the recent editions. This finding confirms hypothesis H4, which says that the early and the latest lists are to a large extent similar in content. I will refer to the 1,520 items as the Core Defining Vocabulary (CDV), because it consists of the most essential vocabulary that has proved stable over time and over all DVs.

The CDV, which is presented in Appendix 1, deserves closer examination because it derives from the intersection of the DVs of several well-established EFL

15. For example, *LDOCE6* uses *Buddhism*, *Catholic*, *Christianity*, *Muslim*, and *Protestant* in the definitions of **Zen**, **friar**, **Jesus**, **harem**, and **pastor**, respectively.

16. Other unlisted grammatical category words such as *idiom*, *consonant*, *pronoun*, *infinitive*, and *syllable* occur in the *OALD9* definitions of **idiomatic**, **closed syllable**, **direct object**, **lemma**, and **feminine rhyme**.

dictionaries. The words it contains are those that a learner is most likely to encounter in definitions. Let us compare it with West's GSL (1953). Interestingly, the CDV shares 92% (1403 items) with the latter list,¹⁷ in spite of the time span between them. This fact points to the stability of this vocabulary over time. Furthermore, the CDV is approximately the size of the vocabulary (i.e. 1,500) that was considered least controversial by the compilers of the GSL (1953), and most frequent in English according to the Faucett-Maki list (1932). Arguably, high-frequency and wide-range words constitute the time-stable core of both the general-service vocabulary and the defining vocabulary.

Analysis of differences between the lists shows certain tendencies. Among 112 items that are on the CDV but missing from the GSL are a few words for food and drink (*onion, potato, vegetable, chocolate, beer, alcohol*), words for vehicles (*aircraft, vehicle*), places (*airport, port*), numbers (*three, four, five, six, seven, eight, nine*), electronic equipment (*television, computer*), verbs (*decorate, establish, remove, obtain, chase, infect*), adjectives (*acceptable, central, frightening, transparent, physical, silly, historical, legal, plastic*), and personal nouns and pronouns (*anyone, everyone, adult, him, himself*). From the lexicographer's perspective, many of these words are useful in defining: most of the nouns mentioned are good candidates for genus terms, whereas most of the personal nouns and pronouns serve as valency indicators.

Compared with Stein's Common Core Vocabulary (CCV), which comprises 2,139 items, the CDV is smaller in size by more than 600 words. Both lists were established on the basis of defining vocabularies from learners' dictionaries. However, while Stein's CCV was compiled from DVs of three dictionaries published in 1995, namely *LDOCE3*, *OALD5*, and *CIDE* (see 2.1), the source material for the CDV constitutes seven lists, from *LDOCE2*, *OALD5*, *CALD1*, *MEDAL2*, *CALD4*, *LDOCE6*, and *OALD9*, published within an extended period of time (between 1987 and 2015). The CDV is more limited in size because fewer lexical items meet the criterion of overlap. An advantage of the methodology used in this study is that the vocabulary thus established is independent of the time of the compilation of the source lists, and is more likely to be resistant to change in the future.

The CDV can be considered as the most essential vocabulary that the learner needs to know in order to understand definitions in contemporary learners' dictionaries. It shares a number of qualities with the CCV. It contains predominantly high-frequency and wide-range words, which are independent of the concepts

17. In this comparison I considered the GSL headwords and the word family members as provided in West (1953). I used word type as the unit of counting, which implied no distinction between part-of-speech homographs; thus, *risk* n. and *risk* v. were counted as a single unit.

being described in definitions. The words are stylistically unmarked and most likely to be encountered by a reader of definitions. This vocabulary can be useful for less proficient learners who would like to reach the advanced level. By acquiring it, the learners can enhance their chances of succeeding in reading comprehension using dictionary definition as aids. Importantly, the CDV contains words most useful in paraphrasing other words, which makes the list useful for learners not only in receptive but also productive tasks, especially when a learner wishes to explain a concept for which they do not know the word (cf. Stein 2002b). For that reason, the list is appropriate for course designers and teachers who work with intermediate students. It can serve as a medium of communication in a vast array of situations and can express complex concepts through paraphrase. The CDV is based on the lexicographers' 30-year practice of paraphrasing a vast range of meanings. Beyond the level of 1,520 items, the dictionaries display differences in the DVs. This is where consistency across the dictionaries is hard to achieve.

Like the CCV, the CDV is as reliable as are the DV lists from which it was extracted. It does not contain phrasal verbs or other multi-word combinations that are sometimes used in the dictionary definitions, because they were not found in all the lists (see Section 4.3.5). Such items are certainly useful in definitions of a limited number of words, but their value for defining numerous other words is questionable. Because of the lack of unanimous agreement among the dictionaries, they were excluded from the CDV.

The CDV has a relatively limited number of transparent derivatives, the meaning of which a learner can easily deduce from those of a base and an affix. For example, the list contains *beautiful*, *careless*, *comfortable*, and *friendly*, but has no *desirable*, *faithful*, *harmless* or *highly*. One of the reasons is that the source lists do not agree on the inclusion of derived words, and another is the policy of affixation in *LDOCE* and, to a lesser extent, *OALD* (see 5.2.1 and 5.2.2). *LDOCE* accords to some of the most frequent and regular affixes the status of independent items in the DV list and allows the definers to use them in definitions to form derivatives that are not on the list. This strategy is restricted only to those derivatives that are semantically predictable and easy for learners to understand. Although the strategy, as it were, diminishes the reliability of the *LDOCE* list, leaving the derivatives unrecorded, the advantage is that the list is short. This is also an advantage of the CDV. The limited size of the CDV (1,520 items) is not without significance for learners, especially at the intermediate level, because it may motivate them in progressing towards proficiency and reduce the learner's psychological barrier to confronting definitions.

Unlike the CCV, the CDV provides no indication of lexical categories but in most cases they can be determined by experienced teachers themselves using their common sense and intuition. Obviously, in the interest of comprehensibility, the

words should be used in prototypical, rather than peripheral, extensional or metaphorical, meanings. This is the criterion to which contemporary dictionaries generally adhere (see 5.2).

A comparison of the CDV against Nation's Base Lists shows that it is composed almost exclusively (99%) of wide-range and high-frequency words belonging to the first four Base Lists. A further inspection of this vocabulary shows that the dictionaries agree on the inclusion of *apple, bread, butter, cake, coffee, egg, flour, juice, meat, milk, onion, potato, rice* (food); *angry, ashamed, grateful, happy, humorous, loyal, pleased* (emotional states and effects); *cat, bird, dog, fish, horse, mouse, sheep* (animals); *arm, ear, eye, finger, foot, head, knee, lung, muscle, nail, neck, stomach, toe* (body parts); *actor, doctor, farmer, lawyer, manager, nurse, officer, soldier* (professions and occupations); *aircraft, bus, bicycle, boat, car, ship* (vehicles); *brother, father, grandfather, grandmother, mother, sister, son* (kinship terms); *airport, bank, church, cinema, college, hospital, restaurant, school, university* (places), *gold, iron, metal, silver* (metals). Many of the words shared by the dictionaries are semantically broad enough to serve as genus words: *animal, feeling, food, material, mixture, organ, organization, place, plant, situation, thing, and vehicle*.

Finally, the CDV contains a group of international words, that is, the English words that have cognates in several European languages, for example, *computer, control, criminal, decorate, decoration, detail, distance, education, and university*. The advantage of using them is that they make definitions easier to understand for those speakers who have similar counterparts in their native languages (Odlin 1989, 77–8; Wingate 2002).

One cannot question the fact that controlled vocabularies developed for different purposes differ in content to some extent. Those designed for trade and international communication are not exactly the same as those for teaching, learning, or defining purposes. Even within the defining activity, the lists will vary depending on the target audience, the level of their language proficiency, the scope of the dictionary, and the direction of its use (receptive or productive). We have also seen differences in the defining vocabularies of different learners' dictionaries. Yet, regardless of purpose, controlled vocabularies share a common lexical core. This core certainly overlaps to a large extent with CDV. Suffice it to say that, as mentioned at the beginning of this section, 92% of the CDV items are general-purpose words that are also found in the dated GSL.

4.3.5 Multiword items

Table 12 presents the proportions of multiword items in the DV lists under study. As we can see, *OALD9* has by far the largest percentage of multiword items (10.0%).

Most of them are semantically opaque phrasal verbs and transparent combinations of verb plus preposition, which have been added after 1995. For example, the list includes the following word combinations involving the verb *come*: *come across*, *come down*, *come from*, *come in*, *come off*, *come on*, *come out*, *come round*, *come to*, *come up*, and *come up to*. The other recent dictionaries contain a significantly smaller proportion of multiword items, which ranges from 0.8% to 2.4%. *LDOCE2* is exceptional, as it records no phrasal verbs.

However, special caution is needed in the interpretation of the above results. As mentioned earlier, the dictionaries do not use the same criteria for assigning entry status in the DVs. There is a mismatch between the inclusion of multiword units in a DV list and their use in definitions. For example, as close inspection of definitions will show (Section 5.2), only the *OALD* list records the semantically transparent combinations *come from* and *send out*, which are used in all the dictionaries. On the other hand, while *LDOCE2* does not list any phrasal verbs, it does use them in its definitions (e.g., *wear out* in the definition of **be hard on**). Furthermore, *few* appears on all the recent lists, but only *CALD4* explicitly permits *a few*. This means that the lists are not exactly true reflections of actual lexicographic practice, but they do indicate changing trends in dictionary policies. We will discuss this issue further in Chapter 5.

Table 12. Multiword items in the DVs of the early and recent editions of the dictionaries.

Dictionary	Phrasal verbs/verbs plus prepositions	Other word combinations	Total
<i>OALD5</i> (1995)	14 (0.4%)	43 (0.7%)	57 (1.6%)
<i>OALD9</i> (2015)	389 (10.0%)	24 (0.6%)	413 (10.6%)
<i>LDOCE2</i> (1987)	0	7 (0.3%)	7 (0.3%)
<i>LDOCE6</i> (2014)	7 (0.4%)	10 (0.5%)	17 (0.8%)
<i>CALD1</i> (1995)	13 (0.3%)	5 (0.1%)	18 (0.5%)
<i>CALD4</i> (2013)	19 (0.7%)	30 (1.2%)	49 (2.0%)
<i>MEDAL1</i> (2002)	24 (0.9%)	37 (1.5%)	61 (2.5%)
<i>MEDAL-2</i> (2007)	28 (1.1%)	32 (1.3%)	60 (2.4%)

4.3.6 Word-building elements

Vocabulary items that dictionaries provide in their DV lists are usually lexemes. They are used in definitions as independent words after inflectional adjustments to fit a context. Some dictionaries list elements smaller than a lexeme, namely derivational affixes (e.g., *-able*, *-er*, *-ish*, *-less*, *-ly*, *non-*, *miss-*) and hyphenated

compound elements (e.g., *-based*, *-boned*, *-edged*, *-length*, *-speaking*), and use them as building blocks for lexemes in definitions. The inclusion of such elements in a DV list expands the defining vocabulary by allowing lexicographers to use unlisted words. Non-DV words are derived from the words already on the list by affixation and compounding. As Table 13 shows, *LDOCE2* relies on affixes more than any other dictionary (2.1%). The other dictionaries either avoid them altogether (*CALD* and *MEDAL*) or permit a handful of them, as is the case in *OALD9*, which lists 6 affixes (0.2%). Hyphenated compound elements appear in *CALD1* (0.7%), but not in the recent *CALD4*. It follows that the DVs in *CALD4* and *MEDAL* are self-contained and self-sufficient, while those in *LDOCE* and, to some extent, *OALD9* allow the lexicographers more latitude in the choice of the defining vocabulary.

Table 13. Word-building elements in the DVs.

Dictionary	Affixes	Hyphenated compound elements
<i>OALD5</i> (1995)	–	–
<i>OALD9</i> (2015)	6 (0.2%)	–
<i>LDOCE2</i> (1987)	49 (2.1%)	–
<i>LDOCE6</i> (2014)	30 (1.4%)	–
<i>CALD1</i> (1995)	–	26 (0.7%)
<i>CALD4</i> (2013)	–	–
<i>MEDAL1</i> (2002)	–	–
<i>MEDAL2</i> (2007)	–	–

4.4 Conclusions

One of the aims of this chapter was to test the hypothesis that the RVLs of definitions in the latest editions of learners’ dictionaries vary considerably (H1). This hypothesis cannot be substantiated by the findings of the analysis, as the vocabulary loads are comparable (3,000–4,000 word families). It should be stressed, however, that only *LDOCE6* and *CALD4* reach the level of 3,000, which is not without significance, as it lends support to the argument that the approach to definition writing (with restricted vs. unrestricted DV) affects the lexical burden of definitions.

Interestingly, the choice of the approach to definition vocabulary has a greater impact on RVL in early editions of learners’ dictionaries than in the most recent editions. For example, the RVL of *OALD4*, which used non-restricted

definitions, decreased by half in *OALD5* with the adoption of a restricted DV. Hornby's approach as used in *ISED* and *OALD4* resulted in definitions being lexically richer and thus more challenging for learners than West's approach in *NMED* and *LDOCE*. Arguably, the former placed greater emphasis on accuracy, while West's opted for comprehensibility (cf. Whitcut 1978, 7; Herbst 1986). However, the difference is no longer so great between the most recent editions of the dictionaries irrespective of the approach to DV.

The hypothesized decrease in the RVL of definitions (H2) has been confirmed for most learners' dictionaries (with multiple editions): *OALD*, *LDOCE*, and *COBUILD*. The above dictionaries also support hypothesis H3, as they show increases in the percentage of potentially useful and known words. In other words, definitions in the recent editions of these dictionaries are lexically lighter and thus easier for learners to read than those in the early editions. The greatest changes in RVL took place in *OALD* up to the 5th edition (1995). No change was found in *CALD* and *MEDAL*, but the values of RVL were already low in the early editions of these dictionaries. Therefore, we conclude that the practice of using the DVs has improved in learners' dictionaries.

The analysis of the DV lists show similar results as that of definitions, as far as hypotheses H1 and H3 are concerned. Contrary to hypothesis H1, the lists show comparably low values of RVL, and in support of hypothesis H3, most of the recent lists rely on a higher percentage of high-frequency words than their early counterparts. Hypothesis H2 has been confirmed for the lists of *CALD* and *LDOCE*, as they show a trend towards lightening the lexical burden.

Another aim of the foregoing analyses was to verify the hypothesis about similarity of content of the DV lists (H4). In support of this claim, we demonstrated that the vast majority (87%) of the vocabulary items shared by the early lists has survived into the recent lists. The dictionaries agree on the use of 1,520 high-frequency and wide-range words, while other words are more or less debatable.

Analysis of native-speaker dictionaries leads us to reject hypothesis H1, which says that the RVLs of definitions in the recent editions of these dictionaries vary considerably. This is because the dictionaries display comparable values of RVL (26,000–28,000 word families). Hypotheses H2 and H3 should also be rejected, because there is no evidence that definitions have become lexically lighter over the years. Nevertheless, definitions in *COD* and *ODE* might be easier to read as compared with the other dictionaries, because they use a greater proportion of words that are very frequent and likely to be known by less proficient users. This is perhaps a sign of the recent trend towards greater accessibility and user-friendliness of definitions, as reported by Atkins and Rundell (2008), as well as Kamiński (2013).

CHAPTER 5

Qualitative analysis

The quantitative analyses presented in Chapter 4 provides insights into the general trends in the policies of controlling definition vocabularies. One of the findings was that the RVL of definitions has decreased in recent learners' dictionaries. This is strong evidence in support of hypothesis H5, that the practice of controlling DVs has improved. The current chapter is intended to give more direct evidence to verify this hypothesis, on the basis of a manual analysis of the development of definitions. The main goal is to see how the dictionaries implemented their policies of DV in successive editions, whether they did it consistently, and if not, for what reasons. These questions would be difficult to deal with using a quantitative method alone. By answering them, we will identify potential problems for users and definers, and determine optimal strategies for controlling the vocabulary of definitions. Section 5.2.7 further explores the use of DVs by looking at digital dictionaries for learners, and Section 5.3 deals with native-speaker dictionaries.

5.1 Research materials and methodology

5.1.1 Materials

The dictionaries under study were the same as those analyzed quantitatively in Chapter 4, but they now also included intermediary editions. Thus, the study focused on the following learners' dictionaries: *OALD* (editions 4–9), *LDOCE* (2–6), *COBUILD* (1–8), *CALD* (1–4), *MEDAL* (1–2), *MWALED*; and the following native-speaker dictionaries: *Chambers* (editions 7, 9, 9new, and 12), *COD* (editions 8, 9, 10, and 12), the *Collegiate* (editions 10–11), *ODE* (editions 1–2), *LDEL*, and *OERD*.

In addition to the paper dictionaries above, I examined several digital dictionaries. The dictionaries selected were available on CD-ROMs or DVDs between 2003 and 2014, and online in December 2019. The dictionaries on optical disks were as follows: *OALD7* and *OALD9*, *LDOCE4* and *LDOCE5*, *CALD3* and *CALD4*, *MEDAL2*, and *COBUILD4*. The study is presented in Section 5.2.7.

5.1.2 Data selection

The study is based on a manual examination of changes implemented in a sample of definitions, and in the DV lists of learners' dictionaries. The latter gave more direct access to the policy of DV, while the former gave evidence of how this policy was executed.

The definition sample consisted of 51 corresponding articles drawn from all the dictionaries and editions under research. The articles for the same headwords were juxtaposed in Tables I–XIV (Appendix 3). For the most part, the sample comprised all the senses of the headwords and subheads. The headwords are listed in Table 14 below. The juxtaposition of the articles allowed the researcher to keep track of changes in definitions in successive editions, and to spot differences among dictionaries. The method enabled the researcher to adopt the lexicographer's perspective and to gain insights into the decision making process involved in the revision of definitions. It also allowed one to see whether the lexicographers' decisions conformed to the dictionary policies as stated in the prefaces or implied by the DV lists. The sample was sufficient for that purpose: the fact that a definer used a particular strategy was sufficient for the researcher to hypothesize on the reasons for that decision. The sample was not subjected to any statistical analysis, so I sometimes found it useful to go beyond the sample range, quoting evidence from other entries. These cases are pointed out in the discussion.

Rather than selecting a run of consecutive entries, I was interested in entries of those words that were difficult to define and interpret. The selection followed three criteria: the potential unfamiliarity of a word for a learner, technical difficulty for the definer, and word category.

The first criterion is strictly related to the frequency and idiomaticity of a lexical unit. The majority of the headwords were morphologically simple low-frequency words, because such words tend to be unfamiliar to learners (Bogaards 1998). They are likely to cause difficulties in decoding because they are rarely encountered in general texts, and have meanings which are unpredictable, or at least cannot be guessed from constituent elements (of which there are none). This is the type of vocabulary that learners look up most often in dictionaries (Béjoint 1981, 217–8)¹. Thus, in the selection, 65% of the headwords are ranked above 10,000 in the ukWaC list (see further below) and 25%, ranked above 50,000 (see Table 14). Idioms constitute another group of lexical units that make contextual

1. In Béjoint's study, foreign learners of English responded to a questionnaire designed to assess their needs and reference skills. When asked "What kinds of words do you look up most often in the dictionary?", 66% of the subjects replied that they never looked up common words, and 47% never looked up structural words (Béjoint 1981, 217).

guesswork almost impossible. Although none of the headwords was an idiom in the strict sense of the word,² the sample included numerous fixed expressions which appeared as subheads. For example, *hard-done-by*, *hard luck*, *hard as nails*, *the hard way*, *hard-headed*, *hard-line*, *hard-nose*, etc., appear in *CALDI*'s long entry for the common adjective **hard**.

Another criterion for the selection of the headwords was their difficulty in defining. Rare words that pertain to specialized and technical fields (e.g., *grout*, *pulsar*, *rattan*, *Terylene*, *wrench*) are especially difficult to define in a restricted vocabulary, as their definitions require the use of other technical or specialized terms (Herbst 1986). Common words, both abstract and concrete, also pose difficulties. As Landau (2001, 167) argues, words like *time* and *life*, in spite of their lexical simplicity and high frequency (both are ranked below 200 in the ukWaC frequency list), express concepts which are "anything but simple".^{3 4 5} The definer is not free from problems when defining *feather*, a simple concrete word referring to a complex thing (Landau, 167). I also added a few natural-kind terms, such as *gherkin* and *parrot*, because they are difficult to paraphrase (Stein 2002b). Therefore, the selection included both concrete and abstract words. In terms of frequency, the selection covered different frequency bands,⁶ and included words occurring in the ukWaC corpus with high, mid, and low frequencies (ranked respectively under 3,000, between 3,000 and 10,000, and above 10,000).

Finally, different word classes were taken into account: nouns, verbs, adjectives, and adverbs, so as not to prioritize a particular defining strategy. This is because different word classes present different levels of difficulty for definers. By and large, verbs are more difficult to define than nouns due to the need to discriminate between a number of different meanings and the need to describe

2. In a restricted sense, idioms are fixed and semantically opaque combination of words (Moon 1998, 4).

3. Samuel Johnson's definition of *network* is a noteworthy example: "any thing reticulated or decussated, at equal distances, with interstices between the intersections" (cited in *LDEL*1984, xi).

4. See also West's discussion of difficulties in defining abstract words within a restricted DV (1935, 12).

5. That simple words are difficult to define is well known to lexicographers. The compilers of the Dictionary of the French Academy published in 1694 observed that it was much easier to define *télescope* than an everyday word like *voir* (Eng. verb *see*) (Miodunka 1989, 29).

6. This division roughly corresponds to Schmitt and Schmitt's classification of vocabulary into high-, mid-, and low-frequency bands (2014). The high-frequency band consists of 3,000 most frequent word-families, and the upper cut-off point for the mid-frequency band is 9,000. Above this level, word-families occur in the language with low frequencies.

complicated relations between verbs and their objects (Landau 2001). In turn, most adjectives cannot be defined by reference to hierarchies of hyponymy, and thus are less amenable to a definition in terms of *genus proximum* and *differentia specifica* (Atkins and Rundell 2008: 445). Consequently, lexicographers often define adjectives in a non-analytical way, using one or more synonymous adjectives. Given the fact that pure synonymy is extremely rare in a language (Cruse 2004), such definitions need to be supported by further information complementing the meaning of the defining words (e.g., register, regional, attitudinal labels). Compared to analytical definitions, synonym definitions when used alone are less effective in conveying the meaning, and they do not “substitute for serious semantic analysis” (Atkins and Rundell 2008, 421). They do not show any semantic relationships which are normally made explicit by analytical phrases (Apresjan 1972, 42). In turn, analytical phrase is not as effective for abstract nouns as it is for concrete nouns (Atkins and Rundell 2008, 415). All the above factors may make definitions difficult for the definer to construct, and for the learner to interpret.

As Table 14 shows, the headwords are dominated by rare words, which are ranked above 10,000 in the ukWaC list. This is justified by the fact that they are most likely to be unfamiliar for learners.

Table 14. Headwords selected for the qualitative research.⁷

Frequency ranges in the ukWaC list	0–3,000	3,001–10,000	10,00–50,000	50,001–above
Headwords	<i>can</i> , <i>n.</i> <i>hard</i> , <i>a.</i> , <i>adv.</i> <i>hire</i> , <i>v.</i> <i>permission</i> , <i>n.</i> <i>potential</i> , <i>a.</i> <i>skill</i> , <i>n.</i>	<i>accuse</i> , <i>v.</i> <i>compulsory</i> <i>distinguished</i> , <i>a.</i> <i>feather</i> , <i>n.</i> <i>hormone</i> , <i>n.</i> <i>insult</i> , <i>v.</i> <i>loom</i> , <i>v.</i> <i>persuade powder</i> , <i>n.</i> <i>ski</i> , <i>n.</i> <i>ski</i> , <i>v.</i> <i>trumpet</i> , <i>n.</i>	<i>abject</i> , <i>a.</i> <i>emba</i> <i>rrassembarrass</i> <i>ed embarrassing</i> <i>embarrassment</i> <i>grout</i> , <i>n.</i> <i>hiccup</i> , <i>n.</i> <i>horde</i> , <i>n.</i> <i>instigate</i> <i>inquisitive loophole</i> , <i>n.</i> <i>parrot</i> , <i>n.</i> <i>pulsar</i> <i>pungent redeem</i> , <i>v.</i> <i>solitude steppe</i> , <i>n.</i> <i>stifle</i> , <i>v.</i> <i>stipend</i> , <i>n.</i> <i>trudge</i> , <i>v.</i>	<i>dimple</i> , <i>n.</i> <i>gherkin</i> <i>hillbilly hobnob</i> , <i>v.</i> <i>inscrutable lopsided</i> , <i>a.</i> <i>obfuscate</i> , <i>v.</i> <i>quandary</i> , <i>n.</i> <i>rancid</i> <i>rattan stingy</i> <i>Terylene</i> , <i>n.</i> <i>wrench</i> , <i>n.</i>
Total:	6	12	20	13

7. Approximately one third of this vocabulary was drawn from the vocabulary selected by MacFarquhar and Richards (1983).

Table 14 classifies headwords according to their frequency ranks in the ukWaC list. It is a lemmatized list compiled from the ukWaC corpus, with lemmas displayed in descending order of frequencies (WaCKy2017). Lemmas that occur most frequently in the corpus occupy top ranks in the list, while rare lemmas are assigned low ranks. Lemmas were provided by the TreeTagger (Ferraresi et al. 2008). The source corpus is one of the largest freely available corpora of English, which consists of two billion word tokens (Ferraresi et al. 2008; Baroni et al. 2009). It is a general-purpose corpus of British English, compiled in 2007 from the World Wide Web by gathering pages from the British domain (ending in .uk). The corpus contains a high proportion of “pre-web” texts from various sources (e.g., recipes, sermons, technical manuals, short stories), which have been converted into electronic form and published online, as well as “web-based genres”, such as blogs, discussion sites, and homepages (Ferraresi et al. 2008; Santini and Sharoff 2007; Gill 2015, 74).

As mentioned earlier in this section, most headwords selected for the study were morphologically simple and of low frequency. Morphology and frequency in combination served as useful criteria for identifying words potentially unfamiliar to learners. Using frequency alone would not be revealing, as numerous rare derivatives are semantically transparent and thus inferable to all but beginning learners. For example, *expertness*, *unhappiness*, *unkindness*, *yellowish* are rare because they are ranked 461,679, 28,885, 134,279, and 38,700 respectively in the ukWaC list (unlike their bases, which are ranked below 3,000). Thus, most of the words qualified for inclusion were not only rare but also morphologically simple.

The analysis of the DV lists required the identification of items added to and deleted from successive lists. This analysis involved two steps: the first one was the computational comparison of the lists using the same R scripts as in Section 4.3.3, and the second one was a manual inspection of the results. The results, in the form of lists of non-overlapping items, were analyzed in order to detect possible directions of revisions. Our interest was in how the changes of the lists affected the actual vocabulary used in definitions.

5.2 Learners’ dictionaries (1987–2015)

5.2.1 LDOCE2 (1987)–LDOCE6 (2014)

The editor of *LDOCE2* revised the DV list extensively, adding 226 new items to and removing 199 items from the previous list. As a result, the list no longer contains “the names of actual places, nationalities, religions, and so on, which are occasionally mentioned in definitions” of *LDOCE2* (*LDOCE2* B 15). Such culture-specific

words appear in definitions without special marking. The decision to exclude them from the list may have been based on the fact that they represent a non-linguistic knowledge of the world, which is different from that which is needed by learners for a good command of English (Herbst 1986, 106–7). Analysis of the deletions indicates that the editor has removed names of religions and their followers (e.g., *Buddhism*, *Buddhist*, *Christianity*, *Christian*), names of weekdays (e.g., *Monday*, *Tuesday*), names of months (e.g., *January*, *February*) and titles (*Mr*, *Mrs*, *Miss*). Other removed words such as *camel*, *ox*, *oxen*, *donkey*, *inn*, *crown*, and *mill* must have been regarded as being of limited use. Further, the deletions include the words that could be considered old-fashioned, formal or literary: *cease*, *coarse*, *conquer*, *content* (*adj.*), *gaiety*, *gay*, *latter*, *rejoice*, and *watchman* (cf. Cowie 2012, 156). The removed items include what Bauer and Nation (1993) classify as fairly regular and infrequent affixes: *-dom*, *en-*, *-ese*, *fore-*, and *-hood*. Their removal must have been justified by the desire to impose further restrictions on the use of derivatives in the definitions. As a result, the total number of affixes has decreased from 55 to 49.

On the other hand, the list has been updated by the addition of *aircraft*, *computer*, and *spacecraft*, which are useful in definitions of technology terms. The majority of new items are derivatives of simple words already on the list, for example, *acceptable*, *admittance*, *agreement*, *amusement*, *argument*, which may have been included to satisfy “a desire to appear more transparent in the use of the [controlled vocabulary]” (Cowie 2012, 156). This strategy, however, runs counter to the policy of affixation, as self-explaining derivatives appear on the list along with their bases and affixes (e.g., *accept*, *-able*, *acceptable*). As it will be shown, the list of the following edition, *LDOCE3*, will restrict the inclusion of both derivatives and affixes.

An example of a definition revised according to the list is that of **hard water** below. To conform to new restrictions, the word *lime*, which is no longer on the *LDOCE2* list, has been replaced with *minerals*. The latter word, in spite of being less precise than the former, occurs in the ukWaC corpus three times as frequently as *lime* (36,000 vs. 12,000, respectively), which reduces the risk of a comprehension problem.

hard water (of water) which contains lime preventing soap from mixing properly with the water (LDOCE1)

hard water (of water) containing minerals that prevent soap from mixing properly with the water (LDOCE2)

With regard to multiword items, the only phrasal verb that appeared in the original edition (i.e. *wrap up*) has been removed from the *LDOCE2* list. This suggests that phrasal verbs are not allowable in definitions, unless they are semantically compositional. However, even though they do not figure on the list, they do

appear in the actual definitions, for example, **be hard on** “to wear (something) out easily or quickly,” which misleadingly suggests that the meaning of *wear out* is compositional. Phrasal verbs can cause problems even for advanced learners. Other phrasal verbs used in the *LDOCE* definitions are *come from* and *send out* (found outside the sample), but their absence from the list is defensible given their semantic transparency:

aura an effect or feeling that seems to surround and come from a person or place
(*LDOCE2*)

discharge to send out electricity or lose stored electrical power (LDOCE2)

None of these phrasal verbs will be recorded in the subsequent lists of *LDOCE*. This practice is different from that of *OALD* (editions 5–9), which allows more or less opaque combinations of words.

According to the editor of *LDOCE2* (B15), “only the most common and ‘central’ meanings of the words in the list have actually been used in definitions”. As in the previous edition, the list appended to the dictionary does not specify the meanings in which the words were used, but it can be assumed that the actual list on which the lexicographers worked was much more explicit about it. However, the evidence from the sampled definitions shows the lexicographers did not always succeed in using words in their central or most common meanings. An example is the adjective *low* in the definition of **abject** (see below), which is used in a non-physical meaning of “being bad or below a certain standard”. Incidentally, *LDOCE2* records this meaning as the fourth sense in the entry for **low**.⁸ The definition as such is not informative to the learner who does not know that meaning. The word may cause problems for less proficient learners, though it is inferable from the basic meaning by metaphorical extension. The definition has been reworded in *LDOCE3* without recourse to this word.

abject 1 (of a condition) as low as possible (LDOCE2)

abject 1 **abject poverty / misery / failure etc** the state of being extremely poor, unhappy, unsuccessful etc (LDOCE3)

Other inconsistencies found in the sample of *LDOCE2* are the words that form restricted collocations. An example is the preposition *on* (*payment*) in the definition of **hire**, which is used in a non-central meaning. Other restricted collocations are *break the law* and *charge someone with*, both in the definition of **accuse**. Nevertheless, they do not seem to present problems for advanced learners.

8. *LDOCE2* orders senses in an entry by giving priority to “the most common or most basic meanings” (F 20).

hire “to get the use of (something) for a special occasion or limited time on payment of a sum of money” (LDOCE2)

accuse to charge (someone) with doing wrong or breaking the law (LDOCE2)

Perhaps more challenging combinations of words are non-compositional phrasal verbs, such as *wear out* in the definition of **be hard on** “to wear (something) out easily or quickly”, and *bring back into (favour)* in the definition of **redeem** “to make free of blame or bring back into favour”. The use of the phrasal verbs is inconsistent with the LDOCE2 list, because none of them appears on the list. On the other hand, to conform to the requirements of the list, the editor has removed the phrasal verb *put into (force)* from the definition of **compulsory**. Nevertheless, such replacements are rather irregular.

compulsory put into force by the law, orders, etc. (LDOCE1)

compulsory which must be done by law, by orders, etc; OBLIGATORY (LDOCE2)

Similar findings are reported by Cowie (2002, 157–8), who notes that while numerous difficult collocations were removed from LDOCE2 (e.g., *in favour of* and *make peace* in the definitions of *recommend* and *reconcile*, respectively), the dictionary persisted in using phrasal verbs and other combinations of words (e.g., *break across*, *break down*, *break into*) in their idiomatic and metaphorical meanings (Cowie 1999a, 157–8). As Cowie (2002, 158) points out, although the use of phrasal verbs was not in line with the policy of restricted vocabulary, the decision to use them can be justified by the fact that they contribute to the naturalness of definitions.

As in LDOCE1, outsiders to the list are typed in small capitals to cross-refer the user to the respective entries. In addition, some, but not all, cross-referenced words are further explained by means of glosses. In some entries the editor has managed to reduce the number of cross-references by replacing them with prescribed words. This can be seen in the definition of **gherkin**, where *kept* has replaced *pickled*.

gherkin a small green vegetable which is usually eaten after being PICKLED in VINEGAR (=kept in a sour liquid); type of CUCUMBER (LDOCE1)

gherkin a small green vegetable (a type of CUCUMBER) which is usually eaten after being kept in VINEGAR (LDOCE2)

Capitalization of non-DV words tends to be limited to potentially difficult words. For example, the derivative *self-conscious* (at **embarrass**) appears in small capitals, signaling to the learner that it is not compositional. However, there are certain groups of non-DV items that may appear in normal type. One such a group comprises self-explaining derivatives, for example *extremely* in the definition of

powder. Another includes fairly opaque derivatives and compounds whose meaning can be gathered from a definition which is close by; for example, a learner can gather the meaning of *skibob*, which is used in the definition of *ski*, by looking at the contiguous entry for **skibob**. Such outsiders are not marked explicitly, which extends the actual defining vocabulary. In this regard, however, *LDOCE* is not an exception, as this convention is used in other dictionaries, for example, *OALD5*, *MEDAL*, and *MWALED*. The success of this strategy rests on the assumption that a learner will pass an eye to an adjacent entry to discover what is missing in the definition being read. Of course, the chance of the learner succeeding in this task rests on their familiarity with the dictionary structure and the lexicographic convention. However, regardless of the use or non-use of explicit marking, the task of dealing with unfamiliar words in definitions is left to the learner.

The DV list specifies allowable lexical categories of the words. As in *LDOCE1*, POS labels are not applied consistently to all the words, which may cause difficulties in the interpretation of the list. Certainly, the lexicographers' task was to use the words only in typical categories. Nevertheless, some changes are noticeable in the *LDOCE2* list. For example, *break*, which was not provided with a label in *LDOCE1*, is now labelled with *v.*, which means that it is allowable only as a verb.

Comparing the first two editions, we notice that the editor has expanded some definitions by adding synonyms to analytical phrases (see **abject** below). This strategy was systematically used in *ISED* and Oxford subsequent editions. Synonyms, many of which are rare, are printed in small capitals as the indication of their non-DV status. The provision of synonyms after an analytical definition is also productively useful, and has a potential of broadening the learner's vocabulary. However, an obvious limitation of this strategy is that it is effective as long as the learner is familiar with this convention and understands the preceding analytical definition.

abject 1 (of condition) as low as possible; deserving great pity ... 2 (esp. of people or behaviour) not deserving respect; showing lack of self-respect.

(*LDOCE1*)

abject 1 (of a condition) as low as possible; pitiful; WRETCHED ... 2 (esp. of people or behaviour) showing lack of self-respect; very HUMBLE

(*LDOCE2*)

From the foregoing analysis, it follows that while the *LDOCE2* editor maintained the overall approach to the DV, certain points remained difficult to deal with, such as the systematic application of restrictions on word meanings, and the desire to strike a balance between defining within the DV and defining in natural language (cf. Cowie 1999a).

In *LDOCE3*, the list underwent further revision. Compared with the previous list, 331 items have been removed and 178 added. Among the additions are words of wide meaning, which could serve as genus terms: *movement*, *apartment*,

attitude, document, equipment, hole, institution, owner, relationship, and statement. There are also combinations of words useful for defining oppositions, such as *as opposed to* and *in spite of*. As for the deletions, a number of them must have been considered redundant for the majority of the definitions, but were useful or even indispensable in a few definitions of semantically related words, where they appear in small capitals. Examples of such words are *elephant* (in the definition of **trumpet** *v.*), *devil* (**black magic**), *fairy* (**Cinderella**), *dot* (**dotted line**), *fox* (**vixen**), *handkerchief* (**hankie**), and *ladder* (**fire escape**). On the other hand, *tiger* is inconsistently printed in normal type in the definitions of **bigcat** and **feline**, in spite of being absent from the list. The editor has imposed restrictions on the variety of English used in definitions, by preferring British spelling variants to their American counterparts. Thus, the list no longer provides *behavior, favor, favorable, neighbor, labor, and color*. The editor seems to have been less concerned with the productive needs of the learner, as she removed a number of low-frequency words that in the previous edition served as cross-referenced synonyms. Examples are *diligent* (under **hard**), *incite* (**instigate**), *obligatory* (**compulsory**), *self-conscious* (**embarrass**), and *wretched* (**abject**). Nevertheless, given the fact that they were an integral part of definitions, the decision to remove them must have triggered by the desire to make definitions more intelligible to learners.

In the *LDOCE3* list phrasal verbs received slightly greater attention than in the earlier editions. While *LDOCE1* contained only one phrasal verb (*wrap up*) and *LDOCE2* had none, the items added to the *LDOCE3* list include 9 phrasal verbs: *deal with, let go of, lie down, look after, look for, look sth up, make into, make up, and pick up* (cf. Xu 2012, 374). This selection will survive in its entirety to *LDOCE4*, only to be reduced to 8 in *LDOCE5* and to 7 in *LDOCE6*. The editor allowed these phrasal verbs perhaps on the assumption that they were most useful for defining purposes. In fact, however, the actual number of phrasal verbs used in the *LDOCE* definitions is larger than the list suggests. For example, the definitions below (outside the sample) illustrate the use of semantically opaque phrasal verbs, *take off, put out, set up, give up, and carry out*, which are all missing from the *LDOCE3* list (and subsequent lists). All these definitions will survive into *LDOCE6* in identical form. Perhaps, the phrasal verbs were not used frequently enough in definitions to merit inclusion in the list. It is not without importance for the length of the list, which is kept short thanks to the omission of items that are useful only in a handful of definitions.

jump jet an aircraft that can take off and land by going straight up and down
(*LDOCE3*)

hydroplane a plane that can take off from and land on water (LDOCE3)

lights-out the time at night when a group of people who are in a school, the army etc must put the lights out and go to sleep (LDOCE3)

pitch a tent to set up a tent or a camp for a short time	(LDOCE3)
cession the act of giving up land, property, or rights ...	(LDOCE3)
carry out to carry out a particular activity or process ...	(LDOCE3)

Some revisions of the *LDOCE3* DV must have been triggered by the criticism of the earlier editions, notably *LDOCE1*. The editor tackled the problem of excessive use of non-transparent derivatives in definitions, which had been raised by Jansen et al. 1987 (see Section 3.2.2.2.), by reducing the total number of allowable affixes almost by half, from 49 to 30 (cf. Xu 2012, 373). The removed affixes include infrequent, though fairly regular, morphemes⁹: *-an*, *-en*, *-ery*, *-ship*, and *-ward*. While such affixes as *-ly*, *-th*, *-ful*, *-less*, and *-y* were permitted, semantically transparent derivatives containing these morphemes such as *eighth*, *respectful*, *harmless*, *hopeless*, and *hairy* were no longer on the list because the learner could easily deduce their meaning from the respective elements. On the other hand, the editor has retained “the forms which are common, or which change their meaning when a prefix or suffix is added” (*LDOCE3* B12). This is in line with Herbst’ suggestion that a DV should include “word formations whose meanings are not easily derivable” (1986, 114). This strategy presumably provided the editor with greater control of non-compositional derivatives.

To help the learner disambiguate word meanings, the editor removed from definitions a number of words used in non-central meanings. For example, under **accuse** three restricted collocations *charge (someone) with*, *do wrong*, and *break the law* are no longer in this definition:

accuse to charge (someone) with doing wrong or breaking the law	(LDOCE2)
accuse to say that someone is guilty of a crime or of doing something bad	(LDOCE3)

The new definition of **accuse** dispenses with the metaphorical use of *break* and the less common nominal use of *wrong*, which dates back to *LDOCE1* (and *NMED1935*). Instead of *wrong*, it uses a descriptive strategy with the semantically unambiguous expression *something bad*, clearly suggesting to the non-native user a typical context of use of *accuse*. Likewise, the editor reformulated the restricted collocation *charge (someone) with* as follows: *to say that someone is guilty of*.

In the new definition of **hire**, we no longer find the preposition *on* (*payment*), which was previously used in a peripheral meaning. Also the complex nominal phrase that follows this preposition (*payment of a sum of money*) has been removed. An advantage of this definition is that it sounds more natural.

9. This classification is after Bauer and Nation (1993), who situate these affixes on the 5th level of transparency (see 4.1.3).

hire to get the use of (something) for a special occasion or a limited time on payment of a sum of money (LDOCE2)

hire to pay money to borrow something for a period of hours or days; LET¹ (LDOCE3)

Another definition worthy of mention is that of **skill**. The introduction of the pronoun *you* enabled the editor to dispense with the dangling participial construction “gained by learning and practice”:

skill special ability to do something well, esp. as gained by learning and practice (LDOCE2)

skill an ability to do something well, especially because you have learned and practised it. (LDOCE3)

Let us now consider the following definitions of **can**:

can a closed metal container in which foods are preserved ... (LDOCE2)

can a metal container in which food or drink is preserved ... (LDOCE3)

The revised definition of **can** is not only more inclusive, thanks to mentioning *food or drink* (rather than just *foods*), but is also more consistent with semantic restrictions. The replacement of the plural *foods* with the more common singular *food* is an improvement in both word form and meaning. The meaning denoted by the plural *foods* is more peripheral (“types of food”) than the one conveyed by the singular *food* (“things that people eat”).

On the other hand, under **feather** the need to control meaning has been outweighed by the need to be precise. The LDOCE3 definition uses a more precise but less typical plural form¹⁰ *hairs* as a replacement for a complex but semantically transparent expression *hair-like material*:

feather any of the many parts of the covering which grows on a bird's body, each of which has a stiff rod-like piece in the middle, with soft hair-like material growing from it on each side (LDOCE2)

feather one of the things that cover a bird's body, consisting of a stem with soft hairs growing on either side (LDOCE3)

The editor of LDOCE4 continued the policy of LDOCE3, but revisions were not as extensive as those in the previous edition. A comparison of the lists shows that 44 vocabulary items have been added and 26 removed. Among the additions

10. The meaning of the plural *hairs* is more peripheral than the one of the singular *hair*; this is also reflected in LDOCE3 by placing the latter sense first and the former second in the entry for **hair**.

one finds relational words, which are typical of the metalanguage of definitions (*relating to, refer*), and the words useful in defining technical terms (*email, tape, software, technology, video, website*). The removed vocabulary items include derivatives such as *fifth, sixth, seventh, ninth, threatening*, which are nevertheless used in many definitions with no indication of their non-DV status, presumably because of their semantic transparency. For example, a query run on the CD-ROM version of the dictionary shows that the word *threatening* occurs in 49 definitions with no typographical differentiation.

The editor continued the policy of disambiguating polysemous words by removing words used in non-central meanings. One example is illustrated in the definition of a **hard/left right**, where the definer used the adjective *big* in place of *sharp* (*turn*):

a **hard/left right** a sharp turn to the left or right (LDOCE3)

a **hard/left right** a big turn to the left or right, for example when you are driving (LDOCE4)

However, a closer analysis of occurrences of *sharp* in the CD-ROM version of LDOCE4 shows that lexicographers used the word in other non-central meanings. The examples below demonstrate the use of *sharp* in the meanings of “shape”, “sound”, “taste”, and “pain”, which correspond respectively to the 13th, 10th, 11th, and 5th sense in the entry for **sharp**. The definitions fall outside the sample range.

angular “having sharp and definite corners”

yelp “a short sharp high cry which a person or an animal makes ...”

acid “having a sharp sour taste”

dull ‘to make something become less sharp or clear’

The above meanings are relatively rare in the LDOCE4 definitions. 16% of the total occurrences (152) of *sharp* represent senses other than the central sense “able to cut easily”. Subsequent editors will persist in using the word in the above meanings, as the definitions cited above will survive down to LDOCE6. A similar finding was reported by Gao (2012, 67), who quotes definitions of **star** from LDOCE1–LDOCE4 to show that in spite of revisions in subsequent editions, the editors kept using certain words (*body, mass, and ball*) in extensional meanings. In this regard, the other learners’ dictionaries are very similar, permitting the use of *sharp* in the above meanings. Apparently, consistency in semantic restrictions is difficult to execute without the risk of definitions being too long or convoluted. Nevertheless, learners can easily disambiguate the words by reading illustrative examples.

Compared to LDOCE4, changes in the LDOCE5 DV are minor, with only two items being added (*look sth up* and *thousandth*), and no items removed (cf.

Xu 2012). More revisions are observed in *LDOCE6*. Analysis of its DV list shows that it has been slightly expanded by the addition of 26 items and the removal of 8 items. The latter includes the phrasal verb *make into*, which is used in definitions (e.g., **adobe** “earth and STRAW that are made into bricks for building houses” *LDOCE6*) without being indicated as an outsider presumably because it is fairly transparent in meaning. The additions to the list include category words for people (e.g., *employee*, *sailor*, *visitor*) and grammatical words (*cannot*, *themselves*, *themselves*). Worth noting is that the decision to include the very rare pronoun *themselves*¹¹ may have been made to satisfy the requirement of political correctness, as it is supposed to ensure a non-sexist language.

Examining the development of definitions over the years, we are witnessing changes that do not result directly from the necessity to impose restrictions on the vocabulary, but rather from the desire to make definitions briefer, more precise, natural, and useful for the user in encoding texts. The changes are noticeable especially in *LDOCE3*. For example, in the definition of **can**, the replacement of the set of semantically narrow words *milk*, *oil*, *waste*, *ashes*, *etc.* with the general word *liquid*, which covers typical contents of a can, has arguably been motivated by concerns for brevity, rather than by the need to conform to changes in the DV.¹²

can 2 ... a usu. round metal container ... used for holding milk, oil, waste, ashes, etc. (*LDOCE2*)

can 3 ... a metal container with a lid that can be removed, used for holding liquid (*LDOCE3*)

Other changes have resulted in definitions being a rich source of productively-useful information. This is the case of definitions revised in the *COBUILD*’s style, for example:

in the can ... (of films) completed and ready for showing (*LDOCE2*)

in the can ... a film that is in the can is complete and ready to be shown (*LDOCE3*)

Although it can be argued that the former definition highlights the context in a visually stronger way than the latter, it fails to show how the expression *in the can* is really used. A learner who is unsure of the use of the idiom needs to find confirmation in illustrative examples. The latter definition has the advantage of illustrating a typical context of use of the idiom, with *film* being the subject, and *in the can*

11. In the ukWaC frequency list, *themselves* has a rank 61,211. By contrast, *himself*, *herself*, and *themselves* are ranked below 3,200.

12. Note that all the words removed, but *ash(es)*, are in both the *LDOCE2* and the *LDOCE3* lists.

the complement of the verb *be*. In this respect, the definition is self-contained. It is also more natural, as it is written in a continuous prose.

Another interesting change is in the definition of **grout**, which mentions two other ingredients (*cement* and *lime*) necessary to make grout. This revision has been dictated by the need for more precision: note that sand and water are insufficient to make this substance.

grout a mixture of sand and water that you spread between TILES when you fix them to a wall (LDOCE5)

grout a mixture of sand, water and CEMENT or LIME that you spread between TILES when you fix them to a wall (LDOCE6)

To sum up, major revisions of *LDOCE* were implemented in the second and the third edition, apparently following criticism of the first edition in the 1980s. The revisions were presumably motivated by the need to update the vocabulary, to keep the list short, and to exercise more control over the use of non-compositional units. Other considerations which seemed to be instrumental in the vocabulary selection regarded its usefulness in definitions. The vocabulary actually used in definitions was not restricted to the words from the list. The editors permitted the use of words not listed in the DV but directly derived from it, which made the policy fairly flexible. Restrictions on the DV had to be implemented along with improvements in other aspects of defining such as precision, brevity, and usefulness of definitions for the learner in production tasks.

5.2.2 *OALD4* (1989) – *OALD9* (2015)

OALD4 (1989) was the last edition to control a definition vocabulary with no explicit restrictions. As Cowie, the editor, explains, one of the reasons for not using a restricted DV was to avoid the risk “of using idiomatic combinations or non-literal meanings of listed words” (Cowie 1999a, 160). Moreover, as Cowie maintains,

It can be argued [...] that simplicity is not the only quality that definitions should possess, and that the editor’s goal, when drafting definitions, should be to achieve lexical and structural simplicity insofar as this can be reconciled with precision, economy, and naturalness. (Cowie 1999a: 160)

Examining the sample, we find that *OALD4* shares the shortcomings of the previous editions. This is manifested by the use of words that are neither frequent nor widely used in the language, such as *despicable*, *pulsate*, *daffodil*, *caustic*, *lavatory*, *vindicate*, *throng*, *suffocate*, *stale*, *spanner*, and *rattan*.¹³ Some of them serve as

13. All these words are ranked above 20,000 in the ukWaC frequency list, and belong to Nation’s Base Lists 7–14.

synonyms complementing analytical phrases, whereby aiding a learner in productive tasks (e.g., **stifle** “feel or make (sb) unable to breathe (easily) because of lack of fresh air; suffocate”). Importantly, however, they are not explicitly marked as synonyms, so a learner may not be aware of their status. In the following edition, this shortcoming will be remedied by highlighting them in capital letters. These words significantly contribute to the heavy vocabulary load of the *OALD4* definitions. Recall that 10,000 word families are needed to cover 98% of the definition text.

In the sample of the *OALD4* definitions, we can easily find idiomatic phrasal verbs (e.g., *reason with*), restricted collocations (e.g., *associate with sb*, *in return for*), and rare words (e.g., *throng*, *caustic*, *vindicate*):

horde very large group (esp of people); huge crowd; throng
pungent (of remarks) sharply critical; biting or caustic
redeem save (sb/sth/oneself) from blame; vindicate
persuade cause sb to do sth by arguing or reasoning with him
hobnob spend time (with sb) in a friendly way; associate (with sb)
hire obtain the use of sth or the services of sb temporarily and esp for a short period of time, in return for payment (OALD4)

Comparing definitions in *OALD4* with those in *OALD3*, one cannot help but notice that simplicity of the defining vocabulary has not received as much attention as their precision. For example, in the *OALD4* definition below, the semantically narrow genus word *achievement* has substituted for the broad word *something*:

a feather in one's cap sth one may justly be proud of (OALD3)
a feather in one's cap an achievement, etc that one can be proud of (OALD4)

The new definitions of **hiccup** and **can** mention extra semantic features, which add to their precision:

hiccup sudden stopping of the breath with a cough-like sound (OALD3)
hiccup sudden involuntary stopping of the breath with a sharp gulp-like sound, often recurring at short intervals ... (OALD4)
can 1 metal container, usu with a lid, for liquids, etc ... (OALD3)
can 1 ... metal or plastic container for holding or carrying liquids ... (OALD4)

Nevertheless, there are definitions that mark a shift away from the formal, literary and rare vocabulary (e.g., *wearily*, *goad*, *perplexity*, *expertly*), though such changes are rather irregular:

trudge walk wearily or heavily (OALD3)
trudge walk slowly or with difficulty because one is tired, on a long journey, etc (OALD4)

instigate incite; goad (sb to do sth); cause (sth) by doing this	(OALD3)
instigate cause (sth) to begin or happen; initiate	(OALD4)
quandary state of doubt or perplexity	(OALD3)
quandary state of not being able to decide what to do	(OALD4)
skill ability to do sth expertly and well	(OALD3)
skill ability to do sth well	(OALD4)

Of all editions of *OALD* published up to 1990, *OALD4* made considerably more improvements in the definition vocabulary, but even greater changes were yet to come.

Compared with *OALD4*, the editor of *OALD5* (1995) went a good deal further in controlling the definition vocabulary. While *OALD4* aimed at simplicity of definitions without any pre-determined constraints, *OALD5* followed the *LDOCE* approach relying on an explicitly listed defining vocabulary. This approach has been in use in *OALD* ever since. The following section focuses on the DV of *OALD*.

5.2.2.1 *The DV of OALD5 (1995) and subsequent editions*

The *OALD5* editor selected words for the DV “principally according to their frequency in the language, as revealed by the corpus [British National Corpus], but also as a ‘core’ vocabulary of real value to students of English.” (*OALD5* vi). More detailed explanations of the selection policy is found in *OALD7*, where the DV received the name “Oxford 3000”. The name indicates the size of the vocabulary, suggesting to the learner that little effort is needed to learn English. According to the publishers, the Oxford 3000 was designed not only for defining purposes but also for learning: “The keywords of the Oxford 3000 have been carefully selected [...] as the words which should receive priority in vocabulary study because of their importance and usefulness” (*OALD7*, R99).

For a word to be granted a place in this vocabulary it had to be:

both frequent and used in a variety of contexts. In addition, the list includes some very important words which happen not to be used frequently, even though they are very familiar to most users of English. These include, for example, words for parts of the body, words used in travel, and words which are useful for explaining what you mean when you do not know the exact word for something.

(*OALD7*, R99)

In other words, the selection is based on three criteria: frequency in English, the distributional range across different types of text, and familiarity. The frequency statistics is determined on the basis of the BNC and the Oxford Corpus Collection (*OALD7* R99). The vocabulary contains a group of words that fall outside the most

frequent and widely used words but which can be useful for productive purposes, especially when paraphrasing other words.

Although the main criteria for the selection of the DV for *OALD* do not seem to have changed over the years, the contents of this vocabulary underwent considerable changes. Let us begin with *OALD5*.

While half of the DV of *OALD5* was shared by the other lists from that period (i.e. *LDOCE2* and *CALD1*) (Table 5), the other half represented marked differences. The long list of nearly 3,500 items had approximately 860 unique items, which were missing from the other lists, for example, *allowance*, *atmospheric*, *bacon*, *batsman*, *bid*, *canvas*, *crawl*, *excess*, *endure*, *lavatory* (instead of *toilet*), *loose-fitting*, *mock*, *myth*, *ore*, *pad*, *regiment*, *spectator*, *straw*, and *suppress*. This is a curious collection of words of doubtful usefulness for definers. According to Allen (1996, 47), the list is a “strange mixture of words”, which may have been “built up, or at least topped up, by the editors as they went along”, adding new words when the need arose.

Compared with *LDOCE2* (1987), which heavily relied on affixation, the inclusion of derivatives in the *OALD5* list (and the subsequent lists) was more restrained. The list had no affixes, but included derivatives formed by means of common affixes. Some of them were hardly self-explaining (e.g., *allowance*, *texture*).

The editor allowed for the use of certain words in definitions without special marking, even though they were not in the list. These exceptions included proper names written with an initial capital letter (e.g., *Jesus*, *America*, and *Islam*),¹⁴ and adjectives describing shades of color (e.g., *yellowish* and *reddish-brown*). In addition, as in *LDOCE*, the editor allowed for the use of words whose meaning could be gathered from a neighboring definition. The words in question were “[r]oot words used in defining derivatives and compounds”, for example “‘bleary’ in the definitions for **blearily** and **bleary-eyed** at the entry for **bleary**”. (*OALD5*, 1417).

The introduction of the restricted DV in *OALD5* had a great impact on the vocabulary actually used in definitions. A few examples illustrate the point. The definition of **hard feelings** no longer relies on *resentment* but instead uses words from the list:

hard feelings no resentment or bitterness (OALD4)

hard feelings no anger or bitterness (BITTER 3) remaining (OALD5)

Likewise, in the definition of **feather**, the expression *fringed structures*, containing an outsider to the DV, has been replaced with *parts*:

14. Proper names were also excluded from the DVs of subsequent editions.

feather any of the many light fringed structures that grow from a bird's skin and cover its body (OALD4)

feather any of the many light parts that grow from a bird's skin and cover its body (OALD5)

The new definition of **abject** dispenses with two rare near-synonyms (*contemptible* and *despicable*):

abject (of people, their actions or behaviour) lacking all pride; contemptible; despicable (OALD4)

abject without any pride or respect for yourself (OALD5)

Sometimes the choice had to be made between two words belonging to the DV. For example, the definer of the new definition of **can** has preferred *metal container* to the polysemous word *tin*. This change may have been motivated by the transparency of *metal container* and/or the fact that *tin* was restricted to British English. In the same entry, the formal word *lavatory* has given way to the neutral and international *toilet*.

can sealed tin in which food or drink is preserved and sold ... **the can** ... lavatory. (OALD4)

can a sealed metal container in which food or drink is preserved and sold ... **the can** ... the toilet (OALD5)

Following *LDOCE*, words that do not belong to the DV are printed in small capitals as an indication of their non-DV status. This typographic convention draws the learner's attention to rare words, such as *cucumber*, *vinegar*, and *daffodil*. However, the dictionary lacks systematicity in observing this convention, for instance, *synthetic* and *cautious* (under **Terylene** and **keep one's powder dry**, respectively) appear in normal type, even though they are not on the DV list.

Like *LDOCE*, *OALD5* restricts the range of allowable meanings of defining words. According to the editor, the words are normally used "in definitions in their main or most common meaning but when this is not the case we show this by indicating the sense number of the word in the dictionary" (*OALD5*, 1417). Sense numbers are in parentheses, sometimes together with the canonical form of the word, as in the following example:

no hard feelings no anger or bitterness (BITTER 3) remaining (OALD5)

The parenthetical information serves as a cross-reference directing the user to the relevant sense. However, there are inconsistencies in this regard. While *bitterness* is marked with a sense number, the word *sharp* in the definition of **hard consonant** is not. Sense indication was a new feature of *OALD*, introduced to aid the learner in finding less central meanings of words. However, how far this strategy

was helpful for learners in comprehending definitions is unknown. As the entries for **trumpet** and **hard** show, sense numbers will no longer be used in *OALD6* and subsequent editions.

Examining the sample, we find instances of deletions of words that do not conform to the restrictions on meaning. The examples below show that the polysemous verb *associate*, which in *OALD4* was used in the meaning of “to spend time”, no longer appears in the *OALD5* definition, though the word as such is allowable.

hobnob ... spend time (with sb) in a friendly way; associate (with sb) (*OALD4*)

hobnob to spend a lot of time with sb, especially sb who is rich and/or famous (*OALD5*)

Incidentally, in the *OALD5* definition, the use of overt pronouns *sb* and *sth* instead of the parenthetical information illustrates a departure from the elliptical style.

The definer of **accuse** achieved greater comprehensibility of the definition by replacing the nominal use of *wrong* (in the part “has done wrong”) with the adverbial one (“has done sth wrong”).

accuse say that sb has done wrong, is guilty (of sth) or has broken the law (*OALD4*)

accuse to say that sb has done sth wrong, is guilty of sth or has broken the law (*OALD5*)

As in *LDOCE*, the criterion of semantic centrality was more difficult to apply in the use of *sharp*, as in the following definition:

hard (of consonants) sounding sharp, not soft (*OALD5*)

It is not unusual to find such meanings in subsequent editions. For example, in the *OALD7* definitions (outside the sample) which are cited below, the word is used in the following senses: “curves”, “frost/wind”, “sounds”, “feeling”, and “flavour/smell”, respectively. The meanings figure as the following senses in the entry for **sharp**: 8, 10, 6, 7, and 9, respectively.

angular having angles or sharp corners

bite a sharp cold feeling

clink to make or cause sth to make a sharp ringing sound ...

neuralgia a sharp pain felt along a nerve, especially in the head or face

tang a strong sharp taste or smell

(*OALD7*)

In the editions subsequent to *OALD5*, the DV lists underwent thoroughgoing revisions. Compared with *OALD5*, *OALD6* has 316 additions and 884 deletions.¹⁵

15. The additions and deletions were counted as types, not tokens. Therefore, the words that were entered twice in the list (e.g., under a different word class or a full form) were counted as

Analysis of the additions shows that almost half of the new vocabulary (42%) is comprised of multiword items. A number of them are grammatical collocations produced by expansion of single words from the list; for example, *on behalf of*, *on sb's behalf* (instead of *behalf*), *in addition*, *in addition to sb*, *in addition to sth* (instead of *addition*), *associated with* (instead of *associate*), *dependent on* (instead of *dependent*), *descended from* (instead of *descend*), *in favour*, *in favour of sb*, *in favour of sth* (along with *favour*), and *in memory of* (along with *memory*). Among the additions one finds a group of lexical collocations (e.g., *fall asleep*, *give birth*, *precious stone*, *pay attention*, *catch fire*), nominal compounds (*chemical element*, *cinema theater*, *movie theater*, *classical music*, *court of law*), phrasal verbs (e.g., *consist of*, *give up*, *look up*, *put on*), and other recurrent combinations of words. These additions were perhaps prompted by the desire to be more transparent in terms of their presentation because most of them are composed of the words already present on the list. Also their inclusion ensures a certain amount of control over allowable combinations of words, restricting non-literal uses. As for deletions, they chiefly consist of single words, especially derivatives, for example, *abnormally*, *absolutely*, *accusation*, *actively*, *arrival*, *appreciation*, *application*, *compensation*, *competitive*, and *enthusiastically*.

The revisions of the DV are reflected in the definition sample. For example, in the definition below the definer has replaced the non-DV word *synthetic* with the synonymous adjective *artificial*:

Terylene a synthetic fabric used esp for making clothes (OALD5)

Terylene a light strong artificial fabric, used for making clothes, etc. (OALD6)

In line with the revised content of the list, some definitions do not use rare and semantically opaque words that in the previous edition figured as cross-referenced synonyms, for example, *self-conscious* (in the definition of **embarrass** below), and *consent* (in the definition of **permission**). These changes were perhaps intended to limit the number of cross-references, but the obvious advantage for learners was that they were not exposed to difficult words.

embarrass to make sb feel awkward, ashamed or SELF-CONSCIOUS (OALD5)

embarrass to make sb feel shy, awkward or ashamed, especially in a social situation (OALD6)

Some decisions involved the choice between two semantically similar words from the DV. For example, in the definition of **obfuscate**, a more familiar and basic word *clear* has been used in place of *confused*, though both words are allowable:

one item. As a result of this methodology, the size of the OALD5 list and the additions as well as deletions in the OALD6 list do not equal the size of the latter list.

obfuscate to make sth confused or difficult to understand, esp deliberately
(OALD5)

obfuscate to make sth less clear and more difficult to understand, usually deliberately
(OALD6)

On the other hand, the editor persisted in using a few difficult combinations of words which did not belong to the DV list. They include semantically opaque phrasal verbs, such as *live on* (under **stipend**), and *take off* in the following definition (outside the sample):

airport a place where planes land and take off ... (OALD6)

Such inconsistencies will also be found in subsequent editions. Apart from the above phrasal verbs, which will survive into OALD9, the definitions below (outside the sample range) illustrate the use of two non-DV items from OALD7: *get up* and *break down*. The latter verb will survive into OALD9. Unsurprisingly, this shows that the actual vocabulary of definitions is larger than is suggested by the list.

knock to hit an opponent so that they cannot get up within a limited time ...
(OALD7 and OALD8)

coeliac disease a disease in which sb cannot DIGEST food (= break it down in their body) because their body is very sensitive to GLUTTEN ...
(OALD7, the non DV items are also found in OALD8)

Some revisions were motivated by the need for greater accuracy and, what Hanks calls, “typification” (1979, 33). Although the new definition of **feather** may seem less informative than the former one, it helps learners easily identify feather by mentioning what is typical of the object. The new definition mentions the important feature *soft*, but dispenses with the definition part that describes the structure of a feather.

feather any of the many light parts that grow from a bird’s skin and cover its body.
A feather has a hollow stem in the centre with fine strands growing out on either side
(OALD5)

feather one of the many soft light parts covering a bird’s body (OALD6)

While revisions of the OALD6 list were dominated by deletions, those in OALD7 center on additions (1,035 items added and 314 removed). The additions make up a large proportion of the list (27%), and include the vocabulary of the developing technology: *cellphone*, *cellular phone*, *click*, *email* (the spelling variant *e-mail* was already on the list), *DVD*, *mobile*, *mobile phone*, and *software*. There is also a small group of more or less¹⁶ regular affixes and combining forms: *anti-*, *ex-*, *-ish*, *mid-*,

16. Affixes such as *anti-*, *ex-*, and *mid-* are fairly regular, while *re-* may be difficult to interpret in numerous complex words such as *recite*, *recollect*, and *recommend*, in which this affix is lexicalized (Bauer and Nation 1993).

non-, *re-*, *self-*, and *-sized*. Most of them, except for *-ish* and *-sized*, will survive into OALD9. Among the additions is *wander*, which has a rather dubious value for the learner as a core defining word. For instance, comparing this verb with *walk* in terms of the criterion of substitution (see Section 2.4), one finds that the former has less potential for substitution than the latter, as it is easier to define *wander* in terms of *walk* rather than the other way around. Surprisingly, however, the examination of all the definitions of OALD7 (on CD-ROM) has revealed that the word is actually not used in the definitions.

The new items on the OALD7 list may suggest that they have not been used in the previous edition. However, the word *software* mentioned above already appeared in the OALD6 definitions without being recorded in the list:

application a program designed to do a particular job; a piece of SOFTWARE
(OALD6)

Note that the word is here printed in small capitals, which justifies its exclusion from the OALD6 list. Likewise, the rare word *crown*, which has been added to the OALD7 DV list, was already used as a cross-reference in the OALD6 definition of **coronation** (outside the sample):

coronation “a ceremony at which a CROWN is formally placed on the head of a new king or queen”
(OALD6)

OALD7 retains both definitions, but the words in question are printed in normal type. This means that the content of the list is to a large extent flexible. There is a group of words used in definitions regardless of whether they are granted a place in the list.

It is interesting to find that the content of the OALD7 list reverted to an earlier state, that of OALD5. This is because nearly 40% of the items removed from OALD6 have been reinstated in OALD7; and likewise, over 40% of the items added to OALD6 are no longer on the OALD7 list. Among the removed items are the names for shades of color, such as *bluish*, *brownish*, *greenish*, *greyish*, *pinkish*, *whitish*, and *yellowish*; as well as numerous multiword items mentioned earlier such as *catch fire*, *chemical element*, *classical music*, *court of law*, *capital city*, *capital letter*, *look up*, *put on*, *care about*, *card game*, *north-east*, and *north-west*. Surprisingly, however, some of these combinations appear in OALD7's definitions without being indicated as non-DV items; for example: *chemical element* (found in the definition of **argon** and many other names for chemical elements), *court of law*, *capital letter*, *capital city*, *look up*, and *put on*. Below are examples of the definitions:

appellate court a court in which people can appeal against decisions made in other courts of law
capitalize to begin a word with a capital letter
metro a large or capital city ...

sky the space above the earth that you can see when you look up, ...

aftershave a liquid with a pleasant smell that men sometimes put on their faces after they shave

What justifies the above deletions is that their meanings are inferable from their constituent elements. Thus, over the years the above combinations of words have not been represented uniformly in the *OALD* lists.

OALD7 was the last edition to provide the DV in the appendix. Since then, the list has been available from the dictionary website (English Oxford Living Dictionaries, 2017). The *OALD7* DV has survived into *OALD8* in virtually identical form. The list has been slightly altered by, *inter alia*, the addition of *afterward* (as a variant of *afterwards*).

The *OALD9* list underwent far more extensive revision, reaching the size of 3876 items. Nearly 1,300 of these items (33%) do not appear in the lists of the other recent dictionaries (Table 6); examples are: *ally*, *chip*, *contemporary*, *cottage*, *conventional*, *curb*, *craft*, *crisp*, *dad*, *diary*, *fellow*, *grandchild*, *guy*, *mall*, *fetch*, *vital*, *yawn*, and *yeah*. Surprisingly, some of them are entered in the list along with semantically close words, for example, *contemporary* along with *present*, *grandchild* with *granddaughter* and *grandson*, and *dad* with *father*. A quick check of the DVD version shows that some of the unique items (e.g., *conventional*, *guy*, *vital*, and *yeah*) never occur in definitions, while others occur in very few definitions; for example, *dad* is used only once, in the contextual part: “the phrase ‘dad’s army’ is ...”. This finding predisposes one to suspect that the *OALD9* list is not a reliable representation of the defining vocabulary.

Comparing the list with the previous one, we find 436 additions and 227 deletions. The vast majority (83%) of the additions are phrasal verbs or verbs plus prepositions. The inclusion of some phrasal verbs (e.g., *account for*, *amount to*, *call off*, *rule out*) raises the question whether they are absolutely indispensable to definitions. By clicking on any of the word combinations, the user is directed to the entry for the corresponding phrasal verb (not just verb), which suggests that all of them are non-compositional. However, using the advanced search function of the CD-ROM version of the dictionary, one finds that only some of the verbs are used in definitions in opaque meanings (e.g., *carry out*, *deal with*, *take off*, and *catch up*¹⁷), whereas others are either used as free combinations of verb plus preposition, for instance *look through* in the definition of **binoculars** and *turn up* in the definition of **hognose snake** (below), or are missing in the definitions (*account for*, *amount to*, *call off*, *look on with*, *rule out*).

17. *Catch up* is used only in the full-sentence definition of **catch up with somebody**.

binoculars an instrument, like two small telescopes fixed together, that makes objects that are far away seem nearer when you look through it (OALD9)
hognose snake an harmless American snake with a nose which is turned up (OALD9)

On close inspection, however, one sees that the last group of words do appear in definitions, but not as word combinations, but as separate words, for example *accounts* and *for* in the definition of **purser** “an officer on a ship who is responsible [...] for the accounts”. This is another piece of evidence for a mismatch between the vocabulary of the OALD9 definitions and that of the Oxford 3000. The surplus words on the latter list do not belong to the actual defining vocabulary.

Interestingly, some of the combinations must have been added in the interest of greater transparency, as they have already been used in the previous editions, but without being granted a place in the DV. For example, OALD7 uses *believe in* (**apostle** “a person who strongly believes in a policy ...”), *arrive at* (**appear** “to arrive at a place”), *break off* (**chip** “to cut or break small pieces off sth with a tool”), and *pick up* (**dialling tone** “the sound that you hear when you pick up a telephone ...”). Other items added include a group of cardinal and ordinal numbers (e.g., *three* and *third*, respectively), which also appear in earlier definitions (e.g., **base-man** “a player who defends first, second or third base” OALD7).

As regards deletions to the OALD9 list, it is interesting that a large proportion of them (78%) constitute additions to the previous list. Examples include variant forms of phrasal verbs (e.g., *pick sth up*, *put sth on*), and restricted collocations (e.g., *take action*, *take advantage of*, *give birth to*, *set fire to*, *make friends*, *take care of*, *in exchange for*). The deletion of the former group can be explained by the presence of their alternative forms on the list (i.e. *pick up*, *put on*, etc.). However, in the latter group, all the collocations mentioned above still appear in the definitions, which is even more surprising that they are not fully compositional. Examples of the definitions are **back off** “to choose not to take action ...”, **grab** “to take advantage of an opportunity to do or have something”, **abort** “to give birth to a child or young animal too early for it to survive”, **put something to the torch** “to set fire to something deliberately”, **overture** “a suggestion or an action by which somebody tries to make friends ...”, **babysit** “to take care of babies ...”, **in favour** “in exchange for another thing” (OALD9). It should be stressed that the above reservations are about inconsistencies in the application of the policy of DV rather than the intelligibility of the vocabulary.

Finally, it is worth noting that in OALD7 to OALD9, the words belonging to the DV received special treatment in both the macro- and microstructure. In order to draw the learner’s attention, the headwords are printed in larger type, with

a key symbol to highlight their special status. In addition, they are provided with extra illustrative examples, as well as extra notes on related words and synonyms. The special treatment of keywords is not unique to *OALD*. All the other EFL dictionaries mark high-frequency words with symbols: *COBUILD8* highlights headwords with blue diamonds, *LDOCE6* does so with red circles, and *MEDAL2*, with red stars.¹⁸ These dictionaries prefer ideographic symbols to alphanumeric ones, presumably to achieve visual clarity. The latter symbols appear in early editions of *LDOCE*. For example, *LDOCE3* (1995) uses S1, S2, and S3 to mark respectively the first, the second, and the third 1,000 most frequent words in spoken English; and W1, W2, and W3 to mark the respective ranges of words in written English (*LDOCE3*, xiii). As for *CALD4*, it indicates word senses with CEFR labels, from A1 to C2 (i.e. from beginner to proficiency), so that the learner can know which meanings should be given priority in learning at a particular level of language competence. In the earlier editions (*CALD2* and *CALD3*), the frequency information system used different labels: E(ssential), I(prover), and A(advanced). Essentials were words that are common, useful, and important for learners to know, improvers help them improve beyond basic English, while advanced words make their English more fluent and natural.

In this section we have seen how the shift in the policy of vocabulary control in *OALD5* imposed radical changes on the definition vocabulary. This is clear not only from the study of the receptive vocabulary load, which declined by half in *OALD5* (see 4.2.1.1), but also from the foregoing investigation. However, the most striking finding is the rather flexible selection of items for the DV list. Although the list underwent extensive revision over the years, especially by the addition of numerous combinations of words, the changes did not always go hand in hand with changes of definitions.

5.2.3 *COBUILD1* (1987) – *COBUILD8* (2014)

COBUILD entered the field of EFL lexicography in the year of publication of the second edition of *LDOCE* (1987). *COBUILD* has a few distinctive characteristics which make it especially suitable for productive use. The dictionary exploits full-sentence definitions consistently throughout the dictionary. The word being defined is embedded in the defining sentence, which brings the learner the benefit of showing a typical grammatical context for use of the word (see Cowie 1999a; Sinclair 1987). As a result, numerous definitions consist of two parts: the subordinate clause (e.g., *if/when*-clause), which provides a typical context of use of the

18. For word frequency information in EFL dictionaries see Li (2010).

headword, and the superordinate clause, which constitutes the definition proper (“If you pot a plant, you put it into a pot filled with earth ...” *COBUILD1*). The sentential style was a significant innovation in EFL lexicography, as it allowed the learner to read definitions like ordinary natural English (*COBUILD1*, viii–xvi). Such definitions are believed to be intelligible to “every user” (*COBUILD1*, xvi). An advantage of *when*-definitions is that they convey to a learner syntactic information more successfully than traditional analytical definitions (Lew and Dziemiątko 2006). An essential part of the subordinate clause is the use of the pronoun *you* as a strategy for verbs with human subjects (e.g., “If you assure someone that something is true, ...”). This strategy serves not only to indicate selection preferences of a verb, but also to imitate the language that ordinary people use when explaining words to a fellow human being (Stein 2002a, 106).¹⁹ The idea is to create a certain degree of intimacy between the dictionary and the learner in order to help them understand definitions. Although the full-sentence style is not entirely novel to English lexicography,²⁰ it has provided an inspiration for other lexicographic traditions.²¹

19. Interestingly, the use of the pronoun *you* in *COBUILD1* did not always achieve a desirable effect. This was the case of words for illegal and immoral acts. For example, a foreign learner can be astonished by the definition of **murder** “If you murder someone, you kill them deliberately and in an unlawful way”, which, in spite of the generic use of *you*, can be interpreted as carrying an implication that the user is a serious offender (Stein 1989: 29), or that murdering is acceptable in the English-speaking communities (Hanks 1987: 194). Other definitions carry similar implications, for example, insult “If you **insult** someone ...” (*COBUILD1*). This definition was improved in *COBUILD2* by exchanging the subject with the object: “If someone insults you ...”, though the new definition is not devoid of a depressing effect on a learner (see Hanks 1987).

20. Full-sentence definitions were occasionally used in some early dictionaries of English, for example: “TRANSMUTATION of Metals [with Alchymists] or the Grand Operation (as they call it) is the Finding of the Philosopher’s Stone” (Bailey’s *Dictionarium Britannicum* 1730; as cited in Rundell 2008b, 197).

21. In Polish lexicography there are a few dictionaries that are either modeled on or influenced by *COBUILD*. *Inny Słownik Języka Polskiego* (A Different Dictionary of the Polish Language) (2000) is a general-purpose dictionary for Polish native speakers that uses systematically full-sentence definitions in the same way as *COBUILD* does. It defines words in a simple and natural language, using the knowledge of an average speaker rather than of a scientist (Bańko et al. 2011; Bańko 2001). Therefore, definitions dispense with scientific terms and synonyms, the latter of which being loosely related with the word being defined. Another native-speaker dictionary influenced by *COBUILD* is Dubisz’s *Uniwersalny słownik języka polskiego* (Universal Dictionary of the Polish Language) (2003) (Bańko 2010). The *COBUILD* defining style has been appreciated by Polish lexicographers compiling dictionaries for school students; two notable examples are *Wieża Babel. Słownik wyrazów obcych nie tylko*

In the area of vocabulary control, *COBUILD1* did not follow in the footsteps of *LDOCE*. However, in the early stage of the *COBUILD* project, the editors intensely debated pros and cons of a restricted DV, in order to make the lexicographers aware of the problem. As Fox, an editor of *COBUILD1*, explains,

It was recognized as vital that the people working on the project were alerted to it from a very early stage so that they would have some knowledge of how words should be explained, and of why that way had been decided upon. Our initial feeling very much leaned towards the belief that any kind of vocabulary control is artificial, that words should be explained in whatever other words are needed. We still believe that this is the correct way to proceed. (Fox 1989, 155)

The back cover of the book declares that the definitions have been written in a “carefully controlled” vocabulary with no explicit restrictions specified. In order to ensure that the language was easy for learners to understand, the editors were asked to “monitor carefully the language they choose to use” in the definitions, and to explain “words in other words that are more frequent than the ones they are defining” (Fox 1989, 156). This task was not entirely intuitive, as the compilers had access to computer statistics showing relative frequencies of words (Fox, 156). Using the computer software, the editors were able to classify words into groups according to frequency and to define words in terms of more frequent ones. This strategy was not applied without reservation. In the case of the most frequent words, which cannot be defined exclusively in terms of more frequent ones, the editors were asked to explain the words “as simply but as accurately as possible” (Fox, 156). Thus, although the definition of *give* “If you give something to someone, you offer it to them as a present” makes use of less frequent words (i.e. *offer* and *present*) than *give*, the definition should successfully convey the meaning of the word being defined to “all but the most beginning learner” (Fox, 156). The editors were also instructed to avoid phrasal verbs, as these may be difficult to decode even for advanced foreign learners. However, phrasal verbs were allowable in certain definitions to avoid structural complexity and confusion on the part of the learner, as in the following example: “If something juts out, it sticks out above or beyond a surface or edge” (Fox, 156).

Perusing the sample, one finds that while the *COBUILD1* definitions are generally easy to follow, the reader may be confronted with difficult words in the definitions of *hard water* and *gherkin*. These definitions contain a heavy load of

dla gimnazjalisty (Tower of Babel. A Dictionary of Borrowed Words, Not Only for Secondary School Pupils) by Radosław Pawelec (1999), and *W kilku słowach. Słownik frazeologiczny języka polskiego* (In a Few Words. A Phraseological Dictionary of Polish Idioms) by Katarzyna Mosiołek-Kłosińska and Anna Ciesielska (2001) (Bańko 2010; Bańko et al. 2011).

low-frequency words: *kettle*, *vinegar*, *cucumber*, *pickle*, and *lather*,²² which were nevertheless difficult to avoid without the risk of losing precision. Indeed, the first three words have survived into *COBUILD8*.

Water that is **hard** contains a lot of lime so that it leaves a whitish coating on kettles and soap does not lather well in it.

A **gherkin** is a small green cucumber which has been pickled in vinegar and is eaten as part of a snack or meal.

Other infrequent words are *stale* (in the definition of **rancid**) and *yawn* (**stifle**), both being ranked above 21,000 in the ukWaC list.

In principle *COBUILD1* was “designed for a wide range of users, including students and teachers who need quite a lot of technical information” (*COBUILD1* xvi). However, given the use of rare words and idiomatic combinations of words (albeit occasional one), the dictionary must have served its function better to the advanced than the intermediate learner. Nevertheless, this claim applies just as well to the other learners’ dictionaries under investigation.²³

Subsequent *COBUILD* editors pursued the same approach to defining vocabulary as implemented in the first edition. They aimed to achieve simplicity and naturalness of definitions with no restrictions on the defining vocabulary. The goal was to write definitions “whenever possible” with the words that were “simpler and more common” than the words being defined (*COBUILD2*, xviii, and subsequent editions, e.g., *COBUILD5*, xi). The language of definitions was “modelled on the way people explain the meanings of words to each other” (*COBUILD5*, vii).

As the dictionary progressed, some potentially difficult words were replaced with more familiar ones. For example, consider *lime* and *lather* in the following definitions:

22. In the ukWaC frequency list, these words occupy ranks ranging from approximately 12,000 (*kettle*) to 54,000 (*lather*).

23. In 1999 Kernerman observed that “[s]o far, the most attention in pedagogical lexicography is consistently devoted to advanced English learner’s dictionaries, which really cater for a minority of users (who are more users than *learners* of English). That may be part of deeply-rooted misconceptions in this field – apparently because of historical and commercial factors rather than didactic or realistic ones – causing a virtual conspiracy of publishers, lexicographers, and teachers. However, few non-native learners of English are ‘advanced’; most of them are plain average, and ever more are on a yet lower level.” (1999: 1; cited in Piotrowski 2002). It seems that little has changed since the turn of the century, as the major EFL dictionaries available in the market specify advanced learners as their chief audience (note that all the dictionaries under study explicitly indicate the audience level by using the word “advanced”).

hard water Water that is **hard** contains a lot of lime so that it leaves a whitish coating on kettles and soap does not lather well in it (COBUILD1)

hard water **Hard** water contains a lot of lime so that it leaves a whitish coating on kettles. (COBUILD2)

hard water **Hard** water contains a lot of calcium compounds that stop soap making bubbles and sometimes appear as a deposit in kettles and baths. (COBUILD3–6)

The revisions above reflect the editor's attempt to reconcile the goal of comprehensibility and precision. Although the COBUILD2 definition dispenses with the precise but rare word *lather*, it misses an important feature (of producing lather) that clearly explicates the meaning of **hard water**. This feature reappears in the COBUILD3 definition, but is expressed in simple²⁴ words (*bubbles, soap*). On the other hand, in place of "whitish coating" the definer has used "deposit" in the more precise technical meaning of the word.

The following definitions illustrate a series of lexical replacements resulting in greater simplicity of the defining vocabulary: "sharp sounds" has substituted for "choking sounds" (in the definition of **hiccups**), "old taste" for "stale taste" (**rancid**), "extremely bad" for "shameful or depressing" (**abject**), and "says you must" for "insist" (**compulsory**):

Hiccups are little choking sounds in your throat caused by a quick jerking movement in your chest. You sometimes get hiccups if you have been eating or drinking too quickly (COBUILD1)

When you have **hiccups**, you make repeated sharp sounds in your throat, often because you have been eating or drinking too quickly. (COBUILD2)

If butter, bacon, or other fatty foods are **rancid**, they have gone bad and taste stale and unpleasant. (COBUILD2)

If butter, bacon, or other oily foods are **rancid**, they have gone bad and taste old and unpleasant. (COBUILD3)

You use **abject** to emphasize that a situation or quality is shameful or depressing (COBUILD2)

You use **abject** to emphasize that a situation or quality is extremely bad. (COBUILD3)

compulsory ... is used to describe something that you have to do or accept because it is a rule or law or because someone in authority insists on it. (COBUILD1)

If something is **compulsory**, you must do it or accept it, because it is the law or because someone in a position of authority says you must. (COBUILD2)

24. The words occupy the ranks below 7,300 in the ukWaC list.

A further advantage of the revised definition of **compulsory** above is the expansion of the potentially difficult²⁵ collocation “in authority” into the more transparent phrase: “in a position of authority”.

In the following definition, *respectable* has substituted for the rarer²⁶ word *dignified*:

If you describe someone as **distinguished**, you mean that they look very noble and dignified. (COBUILD2)

If you describe someone as **distinguished**, you mean that they look very noble and respectable. (COBUILD3)

Some lexical replacements must have been prompted by the need to dispense with technical terms. For example, the COBUILD3 definitions of **hormone** and **pulsar** no longer use formal though precise words: *stimulate* and *emit*, respectively. Note in passing that the definition of **pulsar** has also been simplified by the syntactic paraphrase using the relative clause, “a star that spins very fast”, in place of the complex nominal phrase “a rapidly rotating star”.

A **hormone** is a chemical, usually occurring naturally in your body, that stimulates certain organs of your body. (COBUILD2)

A **hormone** is a chemical, usually occurring naturally in your body, that makes an organ of your body do something (COBUILD3)

A **pulsar** is a rapidly rotating star that cannot be seen but which is known to exist because of the regular radio signals it emits. (COBUILD2)

A **pulsar** is a star that spins very fast and cannot be seen but produces regular radio signals. (COBUILD3)

The informalization of the defining language is also seen in the use of the pronoun *you*, which occurs 30 times more in the sample of COBUILD2 than that of COBUILD1 (i.e. 121 vs. 91, respectively).

As in the other dictionaries, the learner may occasionally encounter words used in less typical meanings, for example, the adjective *low* (in “low opinion”) in the definition of **snob**, and *sharp* (in “sharp smell”) in the definition of **pungent**. Neither of them, however, can be a hindrance to advanced users.

Comprehensibility of definitions was not the only goal of revisions. Another was to delete superfluous information, as in the definition of **ski** below. The new definition has been shortened by removing the part that explains how skiers move

25. The difficulty of the collocation *in authority* lies in the polysemous nature of the preposition *in*, which is used in the peripheral meaning and may therefore be ambiguous to a less proficient learner. This ambiguity has been minimized in the revised version of the definition.

26. *respectable* is ranked approximately 10,000, while *dignified* 20,000 in the ukWaC list.

with sticks. It may have been that the definer removed the part to broaden the definition and apply it to water skiing.

When people **ski**, they move on snow wearing skis and using a pair of long sticks to push themselves along, especially as a sport or a holiday activity.
(COBUILD1)

When people **ski**, they move over snow or water on skis. (COBUILD2)

Another goal was to convey productively-useful information. For example, the improved definition of **quandary** no longer defines the word in terms of the semantically vague genus expression “the state of being”, and instead brings out the meaning using an “if-definition”.

A **quandary** is the state of not being able to decide or think what to do about a situation that you are involved in. (COBUILD1)

If you are in a **quandary**, you have to make a decision but cannot decide what to do. (COBUILD2)

Summing up, the selection and use of the defining vocabulary was refined in COBUILD2 and COBUILD3. In the editions published after 2001, the system of vocabulary control survived largely unchanged.

5.2.4 CALD1 – CALD4 (1995–2013)

The DV for CALD1 was selected according to several principles formulated explicitly in the dictionary. The vocabulary consists of “common words of high frequency”, “easy for learners to understand”, “useful for explaining other words”, and expressing “the same meaning in British and American English” (CALD1, 1702). On the other hand, it excludes words that are old-fashioned, and those that are “often confused with other” English or foreign words (CALD1, 1702).

CALD1 deserves much praise for explicit indication of restrictions on word meanings. Some of the words listed in the DV are provided with parenthetical guidewords, that is, semantic pointers indicating the senses in which they are used in definitions, for example, “*ability* (POWER)”, “*singular* (GRAMMAR)”, and “*sing* (MAKE MUSIC)”. Noteworthy is the fact that CALD1 was the first dictionary to provide the user with such information. Interestingly, while CALD drew the learner’s attention to *central* meanings of allowable words, OALD5 (1995) indicated *non-central* ones. We recall that the latter dictionary indicated sense numbers in definitions (see 5.2.2.1). The two strategies served different purposes: while the former provided lexical guidance independently of a particular situation of dictionary use, the latter served as immediate help to understanding a particular definition. However, it is hard to see to what extent the strategies were useful to a learner, as it is the learner, not the editor, who is the best judge of this information.

In either case, a reader who encounters a comprehension problem in a definition is forced to either infer the meaning from the context or look up the entry for the unknown word. This is perhaps a reason why subsequent editions of *CALD* no longer provided a user with as explicit guidance as *CALD1* did.

The DV varies in linguistic form, from morphologically simple words, derivatives, and compounds, to compound elements. Unlike the other lists, it is structured in a hierarchical way; for example, under a common base form *make*, we find derivatives *maker* and *making*, and a multiword expression *make-up*. Although *CALD1* makes use of the largest DV of all the 1995 editions (Figure 2), the vocabulary appears to be self-sufficient, especially with regard to derivatives. While the list has no affixes, it specifies exactly which derivatives are allowable. This practice stands in contrast to that of *LDOCE*, where the lexicographers were free to use unlisted derivatives after applying word formation rules. The advantage of the *CALD* strategy is that it reduces the risk of using semantically opaque derivatives. On the other hand, the policy towards compounds is more flexible, as the list contains a small group of compound elements (e.g., *-edged*, *-facing*, *-filled*, *-flavoured*, *-haired*, *-hater*, *-leaved*), allowing the lexicographers to use compounds that are not on the list. The majority of the compound elements have regular *-ed* and *-ing* suffixes, presumably to ensure semantic and structural transparency. Nevertheless, it is the generous inclusion of semantically transparent derivatives that is a distinctive characteristic of *CALD1*. This feature, however, will come under close scrutiny in subsequent editions.

A close inspection of the list shows that certain phrasal verbs do not have the same status as others. This is due to the fact that, while a small group of common phrasal verbs (e.g., *care for*, *consist of*, *deal with*, *give up*) appear as main entries in the list, others are hidden under the senses of the relevant base verbs. For example, *look for* is not listed as such in the DV, but is inferred from the sense “SEARCH” under *look*. Thus, the number of allowable phrasal verbs is larger than it may seem at a first glance.

Comparing the listed items with their definitions, we find further inconsistencies. For example, although the phrasal verb *carry out* is missing from the list (nor is it implied by a sense indication), it is used in the definition of **position** (outside the sample):

position A position is also the place where people are sent in order to carry out a course of action (CALD1)

CALD uses the same lexicographic convention as we have observed in the other dictionaries. To compensate for a possible loss of comprehension, it sometimes uses outsiders to the DV (e.g., *spanner*, *brass*, *sin*), highlights them in capital letters, and whenever necessary provides extra parenthetical glosses. A learner can

infer the meaning of an unknown derivative from a contiguous definition or entry (e.g., *embarrassed* in the entry for **embarrassing**). As in *OALD5*, although proper names such as *Britain* and *Christianity* are missing from the list, they are used in definitions (of **stipend** and **redeem**, respectively).

A comparison of the DV of *CALD1* with that of *CALD4* demonstrates that the former vocabulary has become smaller by more than 30% (Figure 2). There are 1,456 deletions, and 193 additions.²⁷ Analyzing deletions, one notices large groups of derivatives pertaining to particular word families, for example, *admirable*, *admirably*, *admirer*, *admiring*, *admiringly*; *acceptability*, *acceptably*, *acceptance*, *accepted*; *childhood*, *childish*, *childishly*, *childishness*. Yet analysis of the entire *CALD4* shows that some of these words appear in definitions in spite of being absent from the list. An example is *admiring*, which is used in the definition of **hagiography** “a very admiring book about ...”, and *rave* “admiring and giving a lot of praise”. Another word removed from the list is *accepted*, which is used in as many as 99 definitions, for instance, in the definitions of **agreeable**, **ambassador**, and **approved**. This is perhaps due to the fact that a learner can infer the meaning of *accepted* from the base *accept*. It appears that the editor reduced the list by removing semantically transparent derivatives, but nevertheless admitted them in definitions on the grounds that they create no comprehension barrier to advanced learners. If this was the editors’ line of argumentation, then the policy of the selection of the DV became more consistent with Herbst’s claim that a DV should contain non-compositional words rather than those that are predictable (Herbst 1986, 114).

Among the deletions are semantically opaque words that rarely occur in the language, such as *pretence*, *excrement*, *utensil*, *edible*, *excrete*, and *savoury*.²⁸ A separate group is made up of specialist names for linguistic categories: *pronominal*, *determiner*, *superlative*, *infinitive*, and *auxiliary*. Although the words were not allowable, the lexicographers must have found it difficult to do without them in a few definitions, for example, *utensil* appears in the definition of **spatula**; and *savoury* in the definitions of **aspic** and **cheese biscuit**. In all these cases the non-DV words are explicitly highlighted with capital letters.

As regards the additions, they include recurring combinations of words that are useful in defining abstract concepts, for example, *related to*, *relating to*, *relations*, *rather than*, *in addition*, *in addition to sth*, *about to*, *based on*, *in exchange for*, *in order to*, *opposed to*; a group of technology terms such as *CD*, *digital*, *email*,

27. The additions and deletions were calculated as types.

28. The ranks of these words in the ukWaC frequency list range from around 17,000 (*edible*) to 41,000 (*excrement*).

internet, mobile phone, video, disk, website, and cell phone; a small group of common phrasal verbs: *feed on, fill in, look for, look after, switch sth on, switch sth off, take up, and wake up*; and a set of words for intermediate geographical directions: *southeast, southeastern, southwest, southwestern, northeast, northeastern, northwest, and northwestern*. Both the deletions and additions account for the significant decrease in the receptive vocabulary load of the DV list, from 12,000 to 5,000 word families (Table 4).

All the editions of *CALD* permit the use of words in more than one meaning. For example, in *CALD1* the polysemous verb *charge* is allowable in three senses: “ask (for money)”, “supply energy”, and “control”. As in the other dictionaries, there are some inconsistencies between the list and definitions. For example, the only permissible sense of *cold* is “low temperature”, but in the definition of *hoarse* below (outside the sample), the word is used in the sense of “illness” with no indication of its non-DV status. This contradiction is no longer present in *CALD4*, which allows for the use of the latter meaning.

hoarse (of a voice) sounding rough, often because the speaker has a sore throat or a cold ... (CALD1)

Compared with *CALD1*, the *CALD4* list offers a more restricted number of senses for some words. For example, *CALD1* lists two allowable senses of *fire*: “flames” and “shoot”, whereas *CALD4*, only the former one. However, this is not necessarily reflected in definitions, as *fire* is used in the sense of “shoot” in the definition of *charge* with no typographical distinction:

charge to put enough explosive into a gun to fire it once (CALD4)

A similar typographical inconsistency is with the word *call*, which is allowable in *CALD1* in two senses: “telephone” and “name”. The former no longer appears in the *CALD4* list, but is used in the definition of *dial*:

dial to operate a phone or make a phone call to someone ... (CALD4)

Examining the sample, one encounters a number of lexical replacements implemented in definitions in order to conform to the requirements of the revised list. In the definitions below, the editor has replaced *sum* with *amount* (at *redeem a coupon*), *oily* with *shiny* (at *powder*), and *anxious* with *nervous* (at *embarrass*).

redeem a coupon/voucher, etc. to exchange a piece of paper representing a particular sum of money for that amount of money or for goods to this value (CALD3)

redeem a coupon/voucher, etc. to exchange a piece of paper representing a particular amount of money for that amount of money or for goods to this value (CALD4)

- powder** a soft dry substance which is spread over the skin of the face, in order to stop the skin from looking oily (CALD3)
- powder** a soft, dry substance that is spread over the skin of the face, in order to stop the skin from looking shiny (CALD4)
- embarrass** to cause someone to feel anxious or uncomfortable (CALD2)
- embarrass** to cause someone to feel nervous, worried or uncomfortable (CALD3)

There has been a certain amount of experimentation in the definition of **gherkin**. In CALD2 the definer has replaced *preserved* with *kept*, only to use *preserved* again in CALD4. The reinstatement of *preserved* is justified by the fact that it is more precise than the other word.

- gherkin** a small type of CUCUMBER (= a long green vegetable) which is often PICKLED (= preserved in a vinegar sauce) (CALD1)
- gherkin** a small type of CUCUMBER (= long thin green vegetable) which is often PICKLED (= kept in vinegar) (CALD2)
- gherkin** a small type of CUCUMBER (= long, thin, green vegetable) that is often PICKLED (= preserved in VINEGAR) (CALD4)

It is worth noting that each of the definitions above uses two parenthetical glosses, which is quite extraordinary for a single definition. This fact merits comment. In spite of being intrusive on reading, the use of glosses finds support in the direct method. Furthermore, they are believed to be highly effective in vocabulary expansion (Nation 2009, 102), though it may not be exactly what the average dictionary user seeks. However, it may be frustrating for a learner to having to go to another entry when a gloss is still insufficient for the understanding of a definition. This is especially the case of technical and natural-kind terms, such as *gherkin*, where the first gloss “long, thin, green vegetable” may refer to a cucumber and a bean, as well as an asparagus.

In line with the revised list of CALD4, the editor uses the noun *sympathy* in place of the adjective *sympathetic*. An advantage of the new definition is that it uses a typical collocation *feel sympathy*, making the definition more telling.

- hard-hearted** If someone is **hard-hearted**, they are not kind or sympathetic. (CALD1)
- hard-hearted** If someone is **hard-hearted**, they are not kind or not able to feel sympathy (CALD4)

The definition of **feather** presents another change implemented in line with the list. Here the adjective “hair-like” has been expanded into the full phrase “like hairs”.

- feather** one of the many very light objects with hair-like material along each side
 ... (CALD1)
feather one of the many soft, light things ... with material like hairs along each
 side (CALD4)

The revision of the definition of **hiccup** demonstrates the strict adherence to the vocabulary restrictions. Here the definer has removed a semantically transparent word *uncontrollable*, which is no longer on the list, and used a more descriptive strategy (*that you make in the throat without wanting to*).

- hiccup** a loud noise made in the throat caused by a sudden uncontrollable tight-
 ening of a muscle just below the chest, usually happening repeatedly over a
 short period (CALD2)
hiccup a loud noise that you make in the throat without wanting to, caused by
 a sudden tightening of a muscle just below the chest and usually happening
 repeatedly (CALD4)

Some changes were motivated by the desire to disambiguate a potentially difficult wording. Let us consider the first definition of **accuse** below. Less proficient readers who are unfamiliar with English punctuation rules may mistakenly interpret the participial construction “*to blame (someone), saying*” by treating *someone* as the subject of *saying*. This interpretation may not necessarily be the one that learners eventually accept as true, given the illustrative examples to which they can refer for help, but the construction may nevertheless confuse them and slow down the reading process. The new definition improves on the former by using uninterrupted prose instead.

- accuse** to blame (someone), saying that they have done something morally
 wrong, illegal, unpleasant or unkind (CALD1)
accuse to say that someone has done something morally wrong, illegal or
 unkind (CALD2)

All in all, a number of improvements were introduced to the *CALD* definitions over the years. Restrictions were imposed on the selection of words, their meaning, and combinations of words. On the whole, they were implemented in line with the revised DV, but exceptions to this policy were numerous, showing that the rigorous execution of the restrictions appears to have been difficult to respect.

5.2.5 MEDAL (2002–2007)

As a newcomer to the field of dictionary publishing, *MEDAL* could take advantage of best practice in lexicography. However, at the beginning of the project it was not by any means certain whether the dictionary should adopt a restricted DV. The issue was discussed thoroughly by the editors, and following consultation

with academic advisors, the decision was made that the advantages still outweigh the disadvantages (thanks to Michael Rundell: personal communication). The DV of *MEDAL1* is comprised of “the most common and basic words in English” (*MEDAL1*, 1677).

Unlike *LDOCE*’s, *MEDAL*’s list was self-contained. As the editor explained, “all the words and forms we use are actually shown in the defining vocabulary list: there are no prefixes or suffixes that can make additional words” (*MEDAL1*, 1677). For example, while *tired* and *tiring* were allowable, *tiredness* was not because it was not on the list. This approach made it clear for lexicographers as to which derivatives were acceptable and which were not. The policy had direct implications for definitions. One of the strategies used by lexicographers to conform to the requirements of the list was by paraphrasing. We have already mentioned the *MEDAL2* definition of **hostel**, which used “at low prices” rather than just “cheaply” because the latter word did not belong to the DV.

The editor adopted a more flexible approach to the inclusion of proper names. The list contains the names of religious followers and religions (e.g., *Christian*, *Christianity*, *Jewish*, *Muslim*, *Islam*, *Protestant*), but other proper names which are missing from the list, such as *Asia* and *Russia*, are also used in definitions.

Compared with the DVs in the other early dictionaries, the percentage of multiword items in the *MEDAL1* DV is somewhat higher (2.5%, Table 12). It is largely due to the inclusion of a number of nominal set expressions (e.g., *American football*, *armed forces*, *classical music*, *head teacher*, *middle-class*, *police officer*) and grammatical collocations (e.g., *according to*, *capable of*, *in fact*, *in order to*, *instead of*, *by means of*, *happen to*). The latter group includes numerous useful linking devices commonly used in definitions. It was the editor’s policy to list explicitly all the allowable multiword units, so that the lexicographers would know what was permissible (thanks to Michael Rundell: personal communication). In this regard, *MEDAL*’s policy is more restrictive than *LDOCE*’s.

The *MEDAL* DV as published in the dictionary fails to specify POS labels for the words, but the list that the lexicographers referred to while writing definitions was more explicit about it (Michael Rundell: personal communication). The DV allowed the lexicographers to use words in more than one word class, and it was generally obvious which class was permissible and which was not. For example, it was clear that *ship* could be used as a noun, but not as a verb in the sense “to ship a new piece of software” (Michael Rundell: personal communication). In more difficult cases, the list specified a permissible word class.

In quantitative terms, changes to the *MEDAL2* list are relatively minor, with 31 items added and 25 removed. One of the goal was to bring the DV up-to-date by adding words useful in describing IT terms (*computing*, *download*, *DVD*, *data*, *link*, *password*, *website*). Among the additions there are also names of months and days missing from the previous list, that is, *August*, *September*, *Wednesday*,

Friday, Saturday, and Thursday. As for the deletions, a relatively large group is comprised of semantically predictable derivatives ending in the suffix *-ness*: *fairness, firmness, laziness, politeness, stiffness, sweetness, and unfairness*. These words must have been replaced with other words or retained only in few definitions (see further below).

As in the earlier dictionaries, *MEDAL* occasionally draws upon words outside the list, especially in definitions of technical concepts such as *grout* and *wrench*. The words are highlighted in small capitals, and are sometimes explained in parentheses. As in the other dictionaries, one may encounter some inconsistencies in capitalization of non-DV words. For example, *consonant*, which has been removed from the *MEDAL2* list, appears in the definitions of **palatal**, **velar**, and **semi-vowel**, but in none of these cases is it printed in capital letters, which misleadingly suggests that the word belongs to the DV. Below is an example definition:

palatal “if you pronounce a sound that is palatal, especially a consonant, you pronounce it by moving your tongue near or against the hard **PALATE**”
(*MEDAL2*)

Likewise, although *hairstyle* has been removed from the list, it is used in the definitions of **crew cut**, **Mohican**, **pageboy** with no typographical distinction. However, neither of the inconsistencies affect the comprehension of the definitions.

The editor permits the use of a few common phrasal verbs which are explicitly listed in the DV, for example, *deal with*, *look for*, *take off*, and *get up*. Outsiders are generally hard to find in definitions. A few exceptions are *carry out*, *use up* and *pay back* (see below), but only the last one is found in the sample. The first two have survived into *MEDAL2*, while *pay back* has been replaced with a more general verb *give* (see **redeem** further below).

redeem to pay back all the money you have borrowed (MEDI1)
execution the process of making a computer run a program or carry out an instruction (MEDI1)
burn off to use up energy or get rid of fat from your body by doing physical activity (MEDI1)

Other combinations of words such as *move over (snow)* and *slide over (snow)* (both under **ski**) invite a different interpretation, as they represent semantically transparent structures of verb plus preposition. By the same token, *arrive at* and *look up* are perfectly acceptable.

Some outsiders refer the user to the entry for a related expression, for example *canopener* in the definition below. Such outsiders are highlighted typographically, unless they refer the learner to a definition close by.

can a closed metal container with round sides, for food or drinks. It is opened with a **can opener**. (MEDI1)

The productive *-er* suffix in (*can*) *opener* is used in the instrumental sense which is less common than the agent one (Bauer and Nation 1993). Nevertheless, the meaning of the expression can be deduced from the definition itself, as the context unequivocally suggests a device for opening a can.

Although the *MEDAL* list does not specify allowable senses, in the explanation we read that the words are used “only in the most basic and central meanings” (*MEDAL1*, 1677). The style guide on which the lexicographers worked did not specify meanings either, and for the most part of the work the definers used common sense (Michael Rundell: personal communication). While the sample shows that the lexicographers generally respected restrictions on meaning, one finds a few instances of restricted collocations such as *take care of*, *advantage over*, and *in return for* in the definitions of **charge**, **smirk**, and **hire**, respectively. As in the other dictionaries under analysis, *MEDAL* uses the highly polysemous word *sharp* in several fairly remote meanings (see definitions below). However, neither of the words and collocations can pose comprehension problems for advanced learners, to whom *MEDAL* is addressed.

pungent a pungent taste or smell is very strong and sharp

stitch a sharp pain in the side of your body ...

hairpin bend a very sharp bend in a road, where the road forms a ‘U’ shape“

slap a sharp hit with the palm of the hand (MEDAL1)

Comparing the definitions between the two editions of *MEDAL*, we find no or little changes. The editor continued to restrict the use of phrasal verbs; for example by removing *get back* and *pay back*:

redeem 4 to get an object back by paying someone the money that they paid you for it, especially when the money was a type of LOAN (MEDAL1)

redeem 4 to get an object again by paying someone the money that they paid you for it, especially when the money was a type of LOAN (MEDAL2)

redeem 4a. to pay back all the money you have borrowed (MEDAL1)

redeem 4a. to give all the money that you have borrowed ... (MEDAL2)

Other revisions aimed at greater simplicity. In the following definition, the editor has replaced *important physical processes* with the general word *things*, which does not seem to have impoverished the definition, as the examples *growth and sexual development* sufficiently explain the word. On the other hand, the expression *animals and plants*, used in place of *your body*, has made the definition more informative.

hormone a natural substance produced by your body that controls important physical processes such as those relating to growth and sexual development (MEDAL1)

hormone a chemical substance produced in animals and plants that controls things such as growth and sexual development (MEDAL2)

By and large, the vocabulary of the *MEDAL* definitions was carefully controlled with the use of long-established strategies. The revisions in *MEDAL2* improved the system further by removing some inconsistencies of the previous edition.

5.2.6 *MWALED* (2008)

MWALED is the first EFL dictionary produced by the Merriam-Webster publisher (*MWALED*, 7a). It is also the most recent book of all the EFL dictionaries under study. The definitions are written in “simple and clear language” (*MWALED*, 13a), but without being restricted by an explicit DV.

On the whole, *MWALED* avoids the use of rare words and non-compositional combinations of words in definitions, with a few exceptions. Some rare words appear in definitions of more specific concepts (e.g., *pickles* and *tiles* in the definition of **gherkin** and **grout**, respectively). A learner occasionally encounters cross-references to rare synonyms, but they are clearly distinguished typographically by small capitals, informing the learner that they serve a different function than the preceding analytical definition.

MWALED follows earlier dictionaries, relying on the user’s ability to navigate between definitions of related words. For example, the reader of the definition **pioneer** “to be a pioneer in the development of (something)” who is not familiar with the noun *pioneer* can easily retrieve the meaning from the neighboring definition of the noun. According to the editor, this convention “helps learners both to expand their vocabularies and to gain a fuller picture of a word’s meaning by approaching it from a slightly different direction.” (*MWALED*, 7a).

Phrasal verbs found in the sample, such as *make up*, *give off*, and *buy back*, are more or less transparent, so they cannot be problematic for advanced learners. Below are a few examples of definitions:

feather any one of the light growths that make up the outer covering of the body of a bird
powder a dry substance made up of very tiny pieces of something
pulsar a type of star that gives off a rapidly repeating series of radio waves
redeem to buy back (something, such as a stock or bond) (*MWALED*)

Likewise, the reader will be in no doubt as to the meaning of transparent combinations of a verb and adverb/preposition such as *curve in*, *curve upward*, *glide over*, and *glide on*:

dimple a small area on a part of a person’s body (such as the cheek or chin) that naturally curves in ... 2: a small area on a surface that curves in

ski one of a pair of long narrow pieces of wood, metal, or plastic that curve upward slightly in front, are attached to shoes, and are used for gliding over snow

ski to move or glide on skis over snow or water (MWALED)

A less advanced learner may occasionally encounter minor obstacles caused by words used in non-central meanings and less typical forms. For example, the definition of **feather** below uses the plural form *growths*, as a genus term for *feather*, to refer to an object that has grown on the body. This meaning is distinct from that normally found in general texts, that is, the process of growing. Nonetheless, the form is likely to make the learner realise that a metaphorical sense has been evoked.

feather any one of the light growths that make up the outer covering of the body of a bird

Another source of distraction for a less advanced learner may be the verb *release*, in the definition of **in the can**.

in the can of a film, recording, etc.: completed and ready to be released (MWALED)

The word *released* is used in the rather rare sense of “making something available to the public”. Nevertheless, although the definition does not make it explicit that the word is used in a fairly specific sense, the learner can quickly grasp the meaning by reading the contextual label: “of a film, recording, etc.”. Yet these lexical traps are only quibbles, and they certainly do not pose serious difficulties for advanced learners. Moreover, the examples demonstrate that a proficient user of the dictionary should be able to refer to various parts of the dictionary entry in order to get a full picture of a word.

It is worth noting, in passing, that MWALED uses a Webster-specific strategy of saving space with a series of relative clauses. For instance, in the definition below, successive clauses (... *are attached to shoes, and are used* ...) relate to a common head noun (*pieces*):

ski one of a pair of long narrow pieces of wood, metal, or plastic that curve upward slightly in front, are attached to shoes, and are used for **gliding** over snow (MWALED)

By employing this strategy, which can incidentally be traced back to *Webster's Third New International* (1961),²⁹ the definition has been made syntactically complex but without detriment to its accessibility.

29. The strategy of using a series of verb phrases was introduced in *Webster's Third New International* (1961) by Philip Gove. It was supposed to ‘achieve precision and objectivity’ (Morton

Summing up, the *MWALED* editor paid close attention to the selection of the definition vocabulary, in spite of the lack of explicitly specified boundaries for this vocabulary. Being new in the field of EFL lexicography, *MWALED* drew on the tradition of earlier dictionaries for learners, and undoubtedly avoided the shortcomings of the early dictionaries.

5.2.7 DV in digital dictionaries

The rise of computer technology has led to the move of the print dictionary to the digital medium. Probably the greatest advantage of the digital medium is that defining words are hyperlinked so that a user can easily check the meaning of an unfamiliar word found in a definition (Lew 2004). This section focuses on how electronic dictionaries facilitate the learner in this task, and how they have adjusted their defining vocabulary systems to the learner's needs. The dictionaries under study were published on CD-ROMs and DVDs between 2003 and 2014, as well as online in December 2019.³⁰ All the recent editions of the paper dictionaries under study are available in digital form, either as desktop applications stored on optical disks, or applications for online use or mobile devices.

A great advantage of the digital medium is that it allows dictionary users to easily navigate from one entry to another. This has important implications for a user's ways of dealing with difficult words in definitions. For example, in the DVD-ROM version of *LDOCE5*, which was the last paper edition available on disk, all words in definitions are hyperlinked. By double-clicking on a defining word, the user is immediately provided with the entry for that word. Importantly, the new entry appears in an extra pop-up window, which is displayed alongside the main entry and on the same screen. This double display is a convenient way of dealing with a comprehension gap, because the user does not lose sight of the main entry, unlike in print dictionaries (Dziemianko 2012b, 320). The pop-up window function is also activated in other dictionaries on DVDs/CDs: *MEDAL2*, *OALD9*, *COBUILD4*, and *CALD3*. Interestingly, however, while this function works in *CALD3*, it does not in *CALD4*. In the latter edition, although words in a definition are hyperlinked allowing immediate access to their entries, the main entry immediately disappears once a user clicks on a word in the definition. Consequently, the reader who decides to check the word meaning may find it hard to keep track of what they have been reading in the main window because it is no longer on the screen.

1994, 87). Strictly speaking, however, the new defining style deviated from standard English, but no doubt, it allowed for greater compression of information (Landau 2009, 221).

30. The dictionaries are listed under References as "Dictionaries on CD-ROM/ DVDs" and "Online dictionaries".

Another problem, which is common to all the digital dictionaries, is that hyperlinks are not sense-sensitive. By clicking on a word in a definition, a user is directed to the entry for this word, rather than to the relevant sense in which the word has been used in the definition. If the sense is not given as the first one, which is often the case, a user has to find it by going through the entire entry. This task may be laborious and discouraging.

It should be stressed that the dictionaries available on optical disks offer definitions copied from their print versions. Some of them highlight non-DV words in the same or a similar way as in print. For example, *LDOCE5*, *MEDAL2*, and *OALD7* use small capitals, while *CALD3* underlines such words. Given the automatic linking of defining words with their entries, one may question the sense of highlighting non-DV words. In turn, *CALD4* and *OALD9* adjust typography to the digital form by using normal type. In line with the paper version, *CALD4* offers the same parenthetical glosses to explain non-DV words, as in the definition below. In addition, all the words used in the *CALD4* definitions are hyperlinked.

gherkin a small type of cucumber (= a long, thin, green vegetable) that is often pickled (= preserved in vinegar) (CALD4)

A brief look at the online versions of the dictionaries (accessed in December 2019), makes it clear that they do not offer the same navigation functionalities as the DVD counterparts. One downside of the online dictionaries is that entries for unfamiliar words do not open in new pop-up windows, so a user sees only one entry at a time. Furthermore, the dictionaries vary with regard to the range of navigation pathways. *OALD* makes use of hyperlinks for all words in definitions, and highlights non-DV words in a different color. Both *MEDAL* and *CALD* use hyperlinks for all non-function words, but do not use color or other typographical features to highlight non-DV words, perhaps for the simple reason that they are practically useless. In *LDOCE*, only non-DV are hyperlinked and their outsider status is indicated by underlining. In turn, the online *COBUILD* links only selected words, and surprisingly does so rather inconsistently. For example, in the entry for **grout**, the common word *spaces* is hyperlinked, but the low-frequency *tiles* is not. Likewise, in the definition of **gherkin**, the common adjective *green* has a link, but less common *cucumber* does not. Finally, the online version of *MWALED* appears to be the least helpful in navigation, for it offers virtually no hyperlinks for defining words. Thus, if a user is unsure of a word in a definition, they will perhaps leave the page to find the word meaning somewhere else.

The move from print to digital media has not lifted vocabulary limitations in definitions. So far the publishers have directed their attention to developing the network of links between defining words and their entries, while copying the definition content from the print versions. This represents only the beginning of

digital lexicography, and admittedly does not exploit the full potential of the electronic medium.

Finally, it is worth noting that in the digital age, the formula for an online dictionary is being redefined. Both English speakers and learners refer to the Internet to solve language problems, and many of them treat the web as if it were the only source of information. It is likely that many users look up words without being actually aware of what type of online dictionary they are using and what type of audience it is addressed to. In fact, they may be unaware that dictionaries for foreign learners are anything different from those for native speakers. Although some online dictionaries offer different definitions for different types of users (e.g., the online Merriam-Webster Dictionary offers definitions for native speakers, learners, specialists, and kids), the choice of the appropriate definition (and dictionary) may not be readily apparent for less competent users, who use English as their mother tongue, or second or foreign language. In view of these considerations, definitions worded in simple English have a potential of being useful for a wider audience.

5.2.8 Summary and conclusion

This chapter traced the development of two approaches to vocabulary control and showed strategies used by lexicographers to ensure intelligibility. West's approach, which relied on the use of a controlled defining vocabulary, dominated the learners' dictionary scene, as it was used in *LDOCE2* – *LDOCE6*, *OALD5* – *OALD9*, *CALD1* – *CALD4*, and *MEDAL1* – *MEDAL2*. Hornby's approach, which used no explicit restrictions, found followers in the early editions of *OALD* (up to the 4th edition), all editions of *COBUILD*, and *MWALED*.

The foregoing analysis provides evidence in support of hypothesis H5 that the practice of controlling a definition vocabulary has improved over the years. All the dictionaries (with multiple editions) made improvements in the DVs regardless of the approach to vocabulary control. Revisions are most noticeable in the editions published immediately prior to or following the turn of the century (i.e. *LDOCE3*, *OALD5*, *COBUILD3*, and *CALD2*). The editors improved on accessibility of their definitions by limiting the use of technical and stylistically marked words, idiomatic and non-literal combinations of words, and words used in less common forms and meanings. They worked towards an ever tighter system of DV by paraphrasing potentially difficult terms and unpacking complex nominal phrases into clauses (cf. Crystal 1986, 77). Definitions were revised in line with revisions of the DV lists, but inconsistencies are observable in all the dictionaries. The editors exceeded the framework of the DVs by using a long-established strategy of using extra words which were not recorded in the lists. Improvements in the policy of

vocabulary control are manifested by the more consistent use of wide-range and high-frequency words in the recent editions (Table 2).

The analysis revealed that the editors aimed to reconcile comprehensibility with other qualities of definitions such as precision, naturalness, and economy of expression. The more recent editions definitely pay more attention to intelligibility of definitions than they used to do, the fact that has been supported by the analysis of RVL (4.2.1.1). This is noticeable not only in dictionaries which inherently prioritize comprehensibility of definitions by using a restricted DV (e.g., *LDOCE*), but also in those which do not give so much attention to that quality (e.g., *COBUILD*). The compromise between comprehensibility and the other qualities of a successful definition was difficult to achieve in the earlier editions. Suffice it to mention imprecise definitions worded within the *LDOCE1* DV, and precise but lexically challenging definitions in *OALD4*. In the period under study, there was some experimentation in the policy of defining vocabulary.

All the dictionaries which use a restricted DV occasionally exceed the framework of this vocabulary by using words from outside the list, and sometimes explaining them with extra glosses. Non-DV items are especially noticeable in definitions of specialist concepts. They are normally highlighted in small capitals to catch the reader's attention, but there are some inconsistencies in typography. The convention of highlighting non-DV items was carried over into the electronic versions of the dictionaries.

Derivatives

It is significant that the dictionaries vary in the amount of freedom given to lexicographers in using non-DV items. Differences are especially seen in the use of derivatives and combinations of words. *CALD4* and *MEDAL2* do not list affixes but instead specify permissible derivatives. In this regard, the lists are self-contained and ensure a relatively high level of control over lexicographers. In contrast, *LDOCE6* and, to a lesser extent, *OALD9* follow a more flexible policy, providing the commonest affixes in the list and permitting the use of non-DV derivatives in definitions. Over the years, this policy has become more restrictive in *LDOCE*. This is presumably due to the fact that in the early editions, the strategy of listing affixes sometimes proved counter-productive, giving rise to complex forms displaying semantic idiosyncrasies. This risk was minimized in *LDOCE3* by the introduction of less transparent forms into the DV, and by a significant reduction of the number of permitted affixes.

Combinations of words

All the dictionaries make it explicit that they use certain combinations of words, but they do it to differing degrees. *OALD* records by far the greatest number of

word combinations in the list, whereas *LDOCE* is the least transparent in the use of such items (Table 12). However, the fact that an item is excluded from a DV list is not tantamount to its non-use in definitions. This fact accounts for differences in the sizes of the lists (Figure 2). Inconsistencies are noticeable in the use of semantically opaque phrasal verbs. For example, *LDOCE6* lists *deal with*, *find out*, *look after*, *look for*, and *make up*, but uses other phrasal verbs not recorded in the DV (e.g., *take off*, *put out*, *set up*, *give up*, *carry out*). Inconsistencies are also found in *OALD8*, which uses *live on* and *take off* with no mention of them in the list, and in *CALD4* and *MEDAL2*, which use *carry out* but do not put it in their lists. It seems that by explicit listing of permitted non-compositional combinations of words, the dictionaries gain a higher level of control over collocations and typical word combinations. An advantage of this strategy is that it reduces the risk of using awkward and unnatural structures.

Restrictions on meaning

In addition to the restrictions on the use of lexical forms, the dictionaries restrict word meanings by allowing only those that are the most central and common. Some dictionaries (e.g., *CALD1*) are more explicit about it, indicating explicitly in which meanings the words are used. While the dictionaries generally adhere to these semantic restrictions, exceptions do occur in all the dictionaries, especially in the case of restricted collocations. This is also inevitably the case of high-frequency words, which are difficult to control because of their extensive polysemy (see Rundell 1998, 319). Such words need to be disambiguated by the reader on the basis of context, which can hardly pose problems for advanced learners.

Given the highly polysemous nature of high-frequency and wide-range words, use of words in less central meanings is unavoidable. For example, *sharp* is used in the dictionaries in a variety of meanings, in the context of taste, smell, pain, bend, movement, music, etc. Although it is hard to assume that all of these meanings are equally central, they are nevertheless closely related, which may justify their use. As West pointed out, for a learner a related meaning is easier to learn than a completely new word (GSL 1953, ix).

5.3 Native-speaker dictionaries (1984–2011)

As we have seen in the preceding chapters, the notion of controlled vocabulary is inherent to learners' lexicography. We also noted (see 3.2) that this notion was not given much consideration in native-speaker dictionaries published up to the 1980s. The situation persisted into more recent dictionaries, but there are notable exceptions. In order to have a closer look at the policies of controlled vocabu-

lary in these dictionaries, we will now look at the prefaces and then compare the definitions.

5.3.1 Analysis

The dictionary prefaces mention positive qualities of the definitions provided. Both *COD* and *NODE* aim to achieve a balance between accuracy and clarity of expression. *NODE* places emphasis on “explaining and describing as well as defining” (vii). In definitions of specialist terms, it avoids over-technical vocabulary but ensures “a high level of technical information and accuracy” (vii). *COD12* offers definitions that are not only “accurate, up to date, and objective” but also accessible (viii). Definitions and other entry parts in *OERD* were designed to be informative and readable to users. *LDEL* emphasizes “accuracy, clarity, and consistency of presentation” of information in definitions (xi).

A noteworthy quality of the *LDEL* definitions is the use of explanatory glosses for abstruse defining terms. In the preface, the editor explains the rationale for this strategy, by pointing to Johnson’s maxim that definitions should use “terms less abstruse than that which is to be explained and such terms cannot always be found”. As the editor goes on, difficulties arise when a lexicographer attempts to define technical, scientific, and other abstruse terms, because their definitions require the use of related words that are probably known only to subject field experts or people with an uncommonly broad vocabulary (*LDEL* xi). Users often face “the problems of complete disorientation” when confronted with definitions “in purely technical terms”, and of “frustration when having to hunt through a dictionary in order to find explanations of purported explanations” (*LDEL* xi).

In order to help the user in overcoming these problems, the editor introduced the strategy of “combining a technical, or abstruse defining term with a brief non-technical gloss” (*LDEL* xi). This strategy, drawn from learners’ dictionaries, was employed in the definitions of semantically complex words and words pertaining to specialist vocabulary (*LDEL* xi). As in learners’ dictionaries, terms are printed in small capitals, and the accompanying glosses are in simple words within parentheses, for example:

hiccup a spasmodic involuntary inhalation of air followed by closure of the
GLOTTIS (opening between the throat and windpipe) and a sudden sharp
sound (*LDEL*)

Some glosses serve only as a brief indication of a typical example (e.g., *beer*) of the category denoted by an unfamiliar word (*beverage*). This is sufficient for a less competent user to identify the meaning of the word:

CAN TIN ... *specif*a tin containing a beverage (eg beer) (*LDEL*)

Sometimes definers provide no glosses, but only highlight potentially unfamiliar terms in small capitals. As in learners' dictionaries, this strategy is intended to help users consult entries for defining terms, but its practical value in a print dictionary is limited, and lies in informing a user that a word has an entry in the dictionary. This, however, may be useful in the case of multiword units, such as *wind instrument* and *brass instrument*, especially for users who are not aware that the dictionary provides the entries for these units:

trumpet a WIND INSTRUMENT consisting of a conical or cylindrical *usu* metal tube, a cup-shaped mouthpiece, and a flared free end; *specif* the highest-pitched of the BRASS INSTRUMENT, that has a cylindrical tube with two turns and finger-operated VALVES to vary the pitch (LDEL)

The use of glosses in the LDEL definitions is not as extensive as one may expect from the use of rare words therein. For example, the rare words *smother* and *muffle* in the definitions of **stifle** are left unexplained. Nevertheless, the dictionary deserves a great deal of credit for making definitions of specialist terms "readily comprehensible without distorting or omitting information" (LDEL xi). This strategy is a workable compromise between the intelligibility and accuracy of definitions.

Noteworthy is the fact that LDEL was published in 1984 and is the earliest of all the dictionaries under consideration. The LDEL strategy of glossing terms is an important step forward in the trend towards making definitions comprehensible to different groups of users. It antedates the provision of alternative definitions for experts and non-experts in NODE (see below), and for native speakers, learners, specialists, and children in the online Merriam-Webster Dictionary. Alternative definitions were postulated by Landau nine years later, who envisaged that electronic dictionaries would provide "the folk definition and a scientifically rigorous definition for every scientific or technical entry" (Landau 1993, 117).

Let us now turn to COD. A manual comparison of successive editions of this dictionary reveals refinements in the defining practice, which might contribute to the greater accessibility of the definitions. Amendments are noticeable in COD10, where the editor has removed a number of rare, formal, and old-fashioned words, such as *solitary*, *indict*, *traverse*, *latent*, *procure*, and *niggardly*. The definitions were carried into COD12.

accuse indict (COD9)

accuse claim that someone has done something wrong (COD10)

hire procure the temporary use of (a thing) for an agreed payment (COD9)

hire obtain the temporary use of in return for payment (COD10)

potential capable of coming into being or action; latent. (COD9)

potential having the capacity to develop into something in the future. (COD10)

solitude the state of being solitary	(COD9)
solitude the state of being alone	(COD10)
stingy niggardly, mean	(COD9)
stingy mean; ungenerous	(COD10)
trudge 1 ... go on foot esp. laboriously. 2 ... traverse (a distance) in this way.	(COD9)
trudge walk slowly and with heavy steps, typically because of exhaustion or harsh conditions.	(COD10)

Underlying these changes is the departure from the policy of “the severest economy of expression” of the early editions of *COD* (*COD1*, iv). Recent editions of this dictionary have expanded concise wording by using ordinary prose. As the editor of *COD12* maintains, “the definitions retain the hallmark of conciseness, although this is balanced by an emphasis on clarity and accessibility, using ordinary modern English to explain technical and complex terms” (*COD12*, p. viii). In fact, the simplification of the definition vocabulary was a gradual process, extended over several editions. Revisions of the language of definitions began in *COD6*, continued in *COD8*, and then intensified greatly in *COD10* (Kamińska 2014). The expansion of the *COD* definitions involved a much freer use of common words, which made them easier for users to interpret.

Changes of the *COD10* definitions are significant to the extent that many of them hardly resemble the definitions from the previous edition (Kamińska 2014). This is due to the fact *COD10* (2002), rather than being the follower of the earlier tradition of this dictionary, is a completely distinct work which drew heavily on *NODE* (1998). These dictionaries share the same background: both are the products of the same publishers, and both were compiled under the same editorship of Judy Pearsall (Kamińska 2014). This may also be a reason for the dictionaries to display identical levels of RVL: 26,000 word families (see 4.2.1.2).

The simplification of the definition vocabulary in *COD* was a systematic and regular process. This has been validated by the results of the quantitative analysis of RVL. It will be recalled that *COD* and its source dictionary, *NODE/ODE*, rely more on frequent and wide-ranging words than the other dictionaries (Table 3), which means that their definitions might be easier to read. This can be illustrated by the definition of **abject**, in which the majority of the defining words are ranked below 1,000 in the ukWaC list, and two words occupy ranks between 3,900 and 6,600. In contrast, the other dictionaries use numerous mid- and low-frequency lexemes; among the latter we find *wretched* (ranked 18,000), *despicable* (33,000), *grovel* (44,000), *servile* (65,000), and *self-abasing* (1,261,321).

- abject** (of something bad) experienced to the maximum degree ... 2 completely without pride or dignity (COD12)
- abject** mean, wretched; worthless; grovelling; base or contemptible; cast away (Chambers12)
- abject** 1 *a of a condition* showing utter hopelessness and degradation; deserving great pity ... **b of people or behaviour** showing lack of self-respect; servile, spiritless ... (LDEL)
- abject** 1 miserable, wretched. 2 degraded, self-abasing, humble. 3 despicable. (OERD)

It is worth noting that while COD12 defines **abject** with analytic phrases, *Chambers12* and *OERD* rely extensively on synonym definitions, using rare and non-basic words. Synonym definitions strengthen the productive side of the dictionaries, though they do not offer an accurate explanation of meaning but merely a gist of it. This is the price the dictionaries pay for the policy of comprehensiveness and inclusiveness. Dictionaries that follow such a policy are generally unwilling to compromise on vocabulary restrictions; if they manage to make concessions, they do it as much as limited entry space allows.

All the dictionaries under consideration occasionally use highly technical, scientific, and formal words in definitions. They are found especially in *Chambers* (e.g., *anthracitic*, *exultantly*, *guttural*, *obdurate*, and *sibilant*) and the *Collegiate* (e.g., *barbule*, *barbicel*, *expiate*, *hamulus*, *loess*, *variegate*, *stentorian*, *expostulation*, and *xerophilous*). The words contribute to the heavy vocabulary load of the definitions. Some of the defining terms used in the *Collegiate* definition of **feather** (i.e. *barbule*, *barbicel*, and *hamulus*) do not create serious obstacles to comprehension, because a reader may easily check their meaning by looking at the adjacent picture of the object being defined, specifying the names of each part of feather. Obscure words also appear in *LDEL* (e.g., *atone*, *expiate*, *cornet*), *COD12* (e.g., *commiseration*, *plosive*), *NODE* and *ODE2* (e.g., *conure*),³¹ but they occur there less frequently than in the other dictionaries.

In definitions requiring the mention of specialist terms, the economy of expression and precision emerge as overriding goals. For example, in *COD12*, the definitions of **hiccup** and **hard consonant** use *diaphragm*, *glottis*, *velar*, and *plosive*:

- hiccup** 1. an involuntary spasm of the diaphragm and respiratory organs, with a sudden closure of the glottis and a characteristic gulping sound. 2 a minor difficulty or setback (COD12)
- hard** (of a consonant) pronounced as a velar plosive (as *c* in *cat*, *g* in *go*) (COD12)

31. All these words are ranked above 50 000 in ukWaC.

All the dictionaries analyzed use Latin terms in definitions of natural-kind concepts. Words like *calamus*, *flava*, *zygodactyl*, *cucumis*, *psittaciformes*, *psittacidae*, *anguria*, *sarracenia*, and *sarraceniaceae* serve as taxonomic information. They are extremely rare in the language: they either rank low in the ukWaC list (e.g., *cucumis* and *sativus* are ranked above 150,000) or are missing from this list (e.g., *anguria*, *zygodactyl*, *sarraceniaceae*). In *COD12*, this taxonomic information is provided in square brackets to show that it has a different status from that of the definition proper:

parrot a mainly tropical bird, typically brightly coloured, with a strong down-curved hooked bill and a raucous voice, some kinds of which are able to mimic human speech. [Order Psittaciformes: many species.] (*COD12*)

LDEL and the *Collegiate* employ a similar strategy, placing taxonomy in parentheses:

rattan a climbing palm (esp of the genera *Calamus* and *Daemonorops*) with very long tough stems (*LDEL*)

rattan a climbing palm (esp. of the genera *Calamus* and *Daemonorops*) with very long tough stems (*Collegiate10*)

Using this strategy, *COD* arguably follows *NODE*. The latter dictionary provides encyclopedic information as separate from the main definition, and prints it in a different typeface, following a bullet point (see below). This presentation of information reflects a classical distinction between the encyclopedic and the linguistic knowledge, and is supposed to satisfy the needs of those users who seek accurate scientific information and those who do not have specialist knowledge to understand it.

parrot a bird, often vividly coloured, with a short downcurved hooked bill ...
 • Order Psittaciformes: numerous species, sometimes all placed in the family Psittacidae. The order also contains the cockatoos, lorries, lovebirds, macaws, conures, and budgerigar. (*NODE*)

Examining definitions in the sample, one finds a few amendments introduced for the sake of accessibility. In *Chambers*, they involve the replacement of formal words *pertaining* and *vessel* with *relating* and *container*, respectively:

hard-core pertaining to a hard core (*Chambers7*)
hard-core relating to a hard core (*Chambers9*)

can a vessel for holding or carrying liquids (*Chambers7*)
can a container for holding or carrying liquids (*Chambers9*)

Other examples, which come from outside the sample, illustrate substitution of neutral words for formal ones:

madcap an exuberantly frolicsome person	(<i>Chambers7</i>)
madcap an exuberantly playful person	(<i>Chambers8</i>)
madam a general term of opprobrium for a woman	(<i>Chambers7</i>)
madam a general term of reproach for a woman	(<i>Chambers8</i>)

Although revisions like those above are rather rare in *Chambers9new*, they are a manifestation of a deliberate editorial policy to remove “obsolete and opaque language from definitions” (*Chambers 9new*, vii). The reason behind this policy is the fact that definitions are “no longer widely understood” (*Chambers 9new*, vii). An example of such a definition may be that of **embarrass**, which uses the idiomatic expression *to put out of countenance* and the rare verb *disconcert*.³² These lexical difficulties have been removed from the revised definition:

embarrass to put out of countenance or disconcert	(<i>Chambers9</i>)
embarrass to cause to feel self-conscious, ashamed or awkward	(<i>Chambers9new</i>)

The definition below shows that a definition can gain clarity by introducing a typical example of the category being defined, rather than by mere paraphrasing:

loophole a means of evasion	(<i>Chambers9</i>)
loophole a means of evasion, eg an ambiguity in a contract, etc.	(<i>Chambers9new</i>)

In the *Collegiate* sample lexical modifications are rather hard to find. One is in the definition of **parrot**, where the semantically basic word *colored* has substituted for the rare *variegated*.³³

parrot any of numerous widely distributed tropical zygodactyl birds [...] that [...] are often crested and brightly variegated ...	(<i>Collegiate10</i>)
parrot 1 : any of numerous widely distributed tropical birds [...] that [...] are often crested and brightly colored ...	(<i>Collegiate11</i>)

In the entry for **gherkin**, the definer has improved the transparency of the definition by removing the anaphoric reference which directed the reader to the “fruit” sense of gherkin. The new definition uses *gherkin* in place of the pronoun *it*. As a result, the reader does not need to infer the relevant information from the contiguous definition.

gherkin a small prickly fruit used for pickling; <i>also</i> : the slender annual vine (<i>Cucumis anguria</i>) of the gourd family that bears it	(<i>Collegiate10</i>)
--	-------------------------

32. The lexeme *disconcert* has a rank 27,571 in the ukWaC list.

33. *variegate* is ranked 96,818 in the ukWaC list.

gherkin a small prickly fruit used for pickling; *also* : the slender annual vine
(*Cucumis anguria*) of the gourd family that bears gherkins (Collegiate11)

ODE2 has introduced similar amendments:

can the quantity of food or drink held by such a container (ODE1)

can the quantity of food or drink held by a can (ODE2)

powder a medicine or drug in this form ... (ODE1)

powder a medicine or drug in the form of powder ... (ODE2)

On the other hand, a few changes found in ODE resulted from the need to make definitions more precise. In the definition of **ski**, the definer has replaced *attached beneath* with *fitted to the underside*, and in the definition of **hiccup**, removed *sound like that of a cough*, using *gulping sound* instead.

hiccup an involuntary spasm of the diaphragm and respiratory organs, with a sudden closure of the glottis and a characteristic sound like that of a cough (NODE)

hiccup an involuntary spasm of the diaphragm and respiratory organs, with a sudden closure of the glottis and a characteristic gulping sound (ODE2)

ski a similar device attached beneath a vehicle or aircraft (NODE)

ski a similar device fitted to the underside of a vehicle (ODE2)

All the dictionaries under study use phrasal verbs, which do not pose problems for native speakers, while making the defining language more natural:

hard calling for stamina and endurance (Collegiate10–11)

instigate bring about or initiate (an action or event) (ODE1–2)

instigate to urge on, incite; to initiate, bring about (Chambers12)

instigate bring about by incitement or persuasion (OERD)

redeem to buy back full possession of; repurchase; *esp* to buy back (something used as security in return for a loan) by repaying the money loaned (LDEL)

redeem oneself make up for one's poor past performance or behaviour (COD12)

Admittedly, native-speaker dictionaries do not impose restrictions on meanings of defining words. Thus, it is not unusual to find words used in non-prototypical meanings, such as *short* in the meaning of not having enough of something, as in **hard up** “short of money” (COD12). One can easily find such uses of words in definitions of peripheral and specialized senses, as in **hard** “having a smooth close napless finish” (Collegiate10), in which the technical expression *napless finish* refers to the surface of textile fabric. Such uses of words can hardly be avoided in dictionaries that aim at comprehensiveness.

Thus, the trend towards greater accessibility of definitions is more noticeable in some native-speaker dictionaries than in others. The trend is justified in the face of global expansion of English. English has dominated the international world of science, media, and higher education (Shaw 2013), with numerous academics, journalists, translators and other professionals speaking English as a second or foreign language. They use English with a high degree of proficiency, but they still need a quick reference when producing or understanding a text.

5.3.2 Conclusions

In spite of the fact that the long tradition of native-speaker dictionaries shows little sign of plain English (see Section 3.2), accessibility has become an important feature in the recent editions of some dictionaries. This is especially noticeable in *COD* and *ODE*, which make their definitions more user-friendly by simplifying defining words, and in *LDEL*, which glosses defining terms. In *Chambers* we can also find instances of vocabulary simplification, but the changes are not as regular as in *COD*.

Any dictionary that attempts to offer a comprehensive coverage of the vocabulary of the language risks some of their definitions being incomprehensible to less competent users. To combat this problem, English lexicography has developed ways of presenting definition content on more than one level, depending on whether a reader is or is not familiar with a given subject. To that end, *LDEL* glosses unfamiliar terms, while *COD* and *ODE* separates encyclopedic information from the definition proper. It is believed that these defining strategies will be useful for all readers, ranging from subject experts to people of all walks of life, who do not have specialist knowledge of the subject.

The foregoing analysis supports hypothesis H5, as recent dictionaries place greater constraints on their definition vocabularies than they used to do. As mentioned earlier, improvements in vocabulary control are most noticeable in *COD*. They can be gleaned from the manual comparison of definitions, and from the quantitative examination of the proportion of most frequent and widely used words (in 4.2.1.2).

CHAPTER 6

The defining vocabulary and the user

This chapter seeks to test hypotheses H6–H11 on the basis of a few experiments conducted on EFL learners. One of the major goals is to determine the extent to which a controlled vocabulary contributes to learners' understanding of definitions. In order to minimize interference from other factors, this study will be performed on made-up definitions. Another goal is to investigate how authentic definitions written with and without a DV are perceived and understood by learners, and whether the approach to definition vocabulary (with or without explicit restrictions) affects the understanding of definitions. In Chapter 4 we have seen that both approaches ensure comparable RVLs of definitions, which may suggest that the comprehension of definitions is unaffected by the use or non-use of explicit restrictions on definition vocabulary. In the current chapter we will further explore how the approach to defining vocabulary relates to the RVL and comprehension of definitions. We will also investigate the influence of other factors on comprehension, such as the proportion of frequent and widely-used words in definitions, definition length, and definition type (analytical vs. synonym). These factors certainly do not exhaust the range of possible variables that may affect learners' performance, but they are most relevant to controlled vocabulary.

6.1 Design of the experiment

The experiment consisted of two tasks: one focused on comprehension of definitions (Task 1), and the other, on their perception (Task 2). In the former, subjects were assigned a multiple choice vocabulary test, in which they guessed the meanings of headwords with definitions as aids. The definitions came from different sources (independent variable); they were drawn from existing dictionaries (*LDOCE6*, *COBUILD8*, *ODE2*, and *Chambers12*) and also included made-up definitions, which were invented by the author. The comprehension scores (dependent variable) were compared with the non-parametric Kruskal-Wallis test and

a post hoc test¹ in order to determine whether the scores were affected by definition source. The effect of several factors (i.e. the value of RVL, the proportion of frequent and widely used words, the proportion of synonym definitions, and definition length) on comprehension score was computed with the Kendall's tau correlation coefficient for non-parametric data and tied ranks.

In Task 2, definitions from the four dictionaries were evaluated for their perceived helpfulness, and the ratings were compared across the dictionaries, using a Friedman ANOVA. The experiment differs from previous studies both in the range of dictionaries examined and the focus of the experiment, which was designed to isolate as far as possible the effect of use of a controlled vocabulary on perception and comprehension of definitions. In one part of the research, the study partially replicates MacFarquhar and Richard's method of eliciting judgments on entire dictionary articles from dictionary users, but the material is confined to definitions.

6.1.1 Subjects

The subjects participating in Task 1 were 215 university students whose native language was Polish. They were first-year undergraduates studying English Philology at the University of Opole (149 students) and at the University of Applied Sciences in Nysa (66 students). Task 2 was administered to the same 215 students plus 35 other students of English whose native language was neither Polish nor English. To reduce the effect of students' different proficiency levels, all the participants were university freshmen. The participants were not subjected to a proficiency test beforehand since there was a risk that it would lead to their fatigue, as the tasks proper were time-consuming and repetitive. However, prior to the experiment, subjects were asked to rate familiarity of the headwords used in Task 1 (6.1.3).

6.1.2 Materials

The comprehension task (Task 1) was a multiple choice test constructed with 20 questions, each about the meaning of a headword. The majority of the 20 headwords belonged to the same set of words selected for the definition sample in Chapter 5: *pungent*, *loophole*, *distinguished*, *hillbilly*, *lopsided*, *hobnob*, *stingy*, *obfuscate*, *loom*, *abject*, *redeem*, *solitude*, and *stifle*. The remaining headwords were as

1. All the statistical tests presented in this chapter were conducted in R. This programming language was also used in the preparation of input data and the randomization of the order in which definitions were displayed on a task sheet (see 6.1.2).

follows: *debauchery*, *invidious*,² *impudent*, *infatuate*, *scrawny*, *peevish*, and *superfluity*. They were all low-frequency words,³ and thus most likely to be unfamiliar to learners. As mentioned earlier, students' familiarity with the words was assessed prior to the task.

Questions 1–10 were accompanied by definitions from one of the four dictionaries: *LDOCE6*, *COBUILD8*, *ODE2*, and *Chambers12*. There were four versions of this task, depending on the dictionary from which the definitions were drawn. The above learners' dictionaries were used because they differ in the approach to defining vocabulary: *LDOCE6* uses a controlled DV while *COBUILD8* does not impose explicit restrictions. As for the two native-speaker dictionaries, a quick comparison of their definitions shows that *Chambers12* tends to use synonym definitions for those lexical items that *ODE2* defines analytically. In addition, as reported in Chapter 4, definitions in the former dictionary display a lower percentage of high-frequency and wide-range words as compared to *ODE2* (77.67% and 79.04, respectively).⁴ These factors may have an effect on the comprehension of their definitions. As for questions 11–20, subjects were supplied with made-up definitions written either with or without a controlled vocabulary.

The success of dictionary consultation may depend, among other things, on the presence and quality of examples illustrating the usage of a word, the learner's familiarity with a lexicographic convention, and the learner's linguistic skills as well as dictionary consultation skills. Therefore, in order to limit the effect of external factors on comprehension score, the multiple choice test was prepared in Polish, the native language of the subjects. Only definitions were worded in English. Furthermore, the subjects were tested with no access to the context of the target words (but see Lew 2004), so comprehension was determined solely by definitions rather than the combined information (definition plus context). Such conditions made the task especially difficult for those participants who used native-speaker dictionaries. It should be recalled that McCreary and Dolezal (1999, 132) demonstrated that learners' performance on a multiple-choice test with access to a native-speaker dictionary was "no better than with just the text that they [were] having trouble comprehending". This suggests that native-speaker dictionaries offer little

2. Words *debaucher* and *invidious* were drawn from the selection of vocabulary analyzed by McCreary and Amacker (2006).

3. Their frequency ranks ranged from 7,065 (*distinguished*) to 1,521,120 (*infatuate*) in the ukWaC frequency list; and their median rank reached 60,945 (IQR = 56,928).

4. These percentages refer to the much larger sample of definitions (used in Chapter 4) than the one used in the current experiment. Thus, the percentages were calculated anew for the present sample of definitions (see 6.2.3).

help for learners in such tasks. Although I adopted less realistic test conditions—as the definitions were read in isolation from context—the advantage of this methodology was that the comprehension score was independent of context. Likewise, subjects had no access to other information types normally offered in dictionary entries.

A higher level of precaution was applied to made-up definitions. In order to isolate the effect of the use of a restricted DV, the syntactic pattern of invented definitions was kept simple. The pattern used in controlled and uncontrolled definitions was identical or very similar (see examples below). In this way, the effect of structural complexity was largely reduced, and in some cases eliminated completely. Made-up definitions were created by the author on the basis of existing dictionary definitions (including *LDOCE6*, *CALD4*, and *OED*).

impudent showing no respect to other people
impudent showing unblushing insolence to other people

infatuate to make a person very silly
infatuate to make a person utterly fatuous

stifle to try not to express a feeling
stifle to try to smother a feeling

The selection of definitions was limited usually to a single relevant sense. The multiple choice test had five choices, but only one of them was correct. To minimize random guessing, I added a “nie wiem” [“I don’t know”] option for choice “e”. Incorrect options were distractors. Some of them were designed using false friends; for example, I used the Polish word *ilustrowany* in option “b” (see question 3 below) because it is similar in form to one of the defining words in the *Chambers* definition “illustrious, eminent”.

Other incorrect options were constructed following a kind of *kidrule* strategy. All the same, the test hinged on the understanding of the definitions.

3. Distinguished oznacza:

- a. poważany
- b. ilustrowany
- c. apodyktyczny
- d. mało charakterystyczny
- e. nie wiem

distinguished illustrious, eminent

The subjects for Task 2 were presented with 49 headwords and their definitions, selected from the same four dictionaries as in Task 1. They were the same headwords as those used in the definition sample in Chapter 5, except for *embarrassing*

and *Terylene* because their definitions were missing in at least one of the dictionaries. In order to limit participant fatigue, each student rated half of the entire set of definitions. There were two versions of the task: one with 25, and another with 24 target headwords and their definitions. The difference in the number of headwords between the two versions was negligible, and did not affect students' median ratings. Both versions contained definitions from all the four dictionaries (see Appendix 2). Altogether, there were 196 different definitions to be tested.

Definitions for each headword were printed on a task sheet side by side. As in MacFarquhar and Richards' study (1983), they were ordered randomly with no indication of the source in order to eliminate the risk that students' ratings are affected by their knowledge of the source.

abject (adj.)	abject (adj.)	abject (adj.)	abject (adj.)
the state of being extremely poor, unhappy, unsuccessful etc	(of a situation or condition) extremely unpleasant and degrading	You use abject to emphasize that a situation or quality is extremely bad	mean, wretched
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

In order to avoid confusion on the part of the respondents, only definitions representing the same or similar meaning were selected for comparison. Thus, the selection of senses was generally restricted to a single sense, but in the case of a few definitions from *Chambers12* and *ODE2*, it was necessary to quote two senses to permit a fair comparison. For example, two senses of **obfuscate** "to obscure; to confuse or bewilder" in *Chambers12* were roughly comparable in semantic content with the *ODE2* sense: "make obscure, unclear, or unintelligible".

6.1.3 Procedure

Before the experiment, the subjects participated in a 20-minute orientation session, during which they were instructed what a dictionary definition is and how to read it. The subjects were informed that the experiment tested definitions, and not the subjects. Prior to the experiment, the students were asked to rate familiarity of each headword, on a scale of 1 to 5, from least to most familiar (see "Word Familiarity Page" in Appendix 2). The relationship between subjects' ratings and their comprehension scores was computed with a non-parametric Spearman's rank correlation coefficient. In the comprehension task, the subjects were instructed to read a definition first, and then indicate one of the options that best corresponded to the meaning of the headword. For each correct answer, they received 1 point.

Thus, in this task they could obtain a total of 20 points, with 10 points for questions 1–10 (authentic definitions) and 10 points for questions 11–20 (made-up definitions). The perception task was administered upon the completion of the comprehension task in order to eliminate the risk of learning the meanings of the target words. Two versions of the perception task, containing either 24 or 25 headwords and their definitions, were distributed to the subjects with approximately equal frequency: 127 students received version 1, and 123 students, version 2. The students were asked to evaluate helpfulness of definitions using a rating scale of 1 to 4; with 1 meaning that a definition was unhelpful, and 4, meaning that it was very helpful. The scale was wide enough to include intermediate judgements.

All the instructions in the answer sheets were written in Polish (L1 of the participants in the comprehension task). However, given that 35 students taking part in the perception task were non-Polish, they were given oral instructions in English. There was no time limit for completion, but most students returned their answer sheets within 50 – 60 minutes. The experiment was conducted at the two universities in 2018 and 2019 (both in October), following a pilot study⁵ in May 2018 at the University of Applied Sciences, Nysa.

6.2 Results

The results of the experiment are presented below separately for comprehension of dictionary definitions (6.2.1.), comprehension of made-up definitions (6.2.2.), the influence of three factors on comprehension (6.2.3), and perception of definitions (6.2.4.). A summary of the results is given in Section 6.2.5.

6.2.1 Comprehension of dictionary definitions

The students' ratings of their initial familiarity with the headwords are presented in Table 15. The results are summarized with medians because the data are ordinal. The table shows the number and percentage of students with a particular median of familiarity (from 1 to 5). The table was produced by computing the median of ratings for each student, and then counting the number of students with a particular median. As shown in the table, 64% of the students gave a median rating of 1, and 22% gave a median rating of 2. This means that as many as 86% of the subjects did not know the words.

5. The results of the pilot study were excluded from the final results.

However, what subjects claimed to know did not correspond well to how they actually performed. Analysis of the relationship between subjects' median ratings and their comprehension scores with a non-parametric Spearman's rank correlation coefficient shows that the strength of this relationship was negligible. The ratings were not correlated significantly with the scores (r close to 0, $p < .01$).⁶ For example, 28% (38) of the subjects with low familiarity ratings (median rating = 1) scored as many as 8–10 points in the comprehension test. Even among the students who claimed to know the headwords very well (median rating = 4 or 5), there were those (3) who performed poorly on the test by scoring only 4 points.

Table 15. Students' familiarity with headwords.

Median rating	1	2	3	4	5
Number of subjects	137 (64%)	48 (22%)	19 (9%)	4 (2%)	7 (3%)

Table 16 and Figure 4 show the results of the comprehension task on headwords 1–10, with different dictionary definitions as aids. The table provides median comprehension scores for the dictionaries, and the values of dispersion of the data (interquartile range). The table indicates that the subjects using *LDOCE6* and *COBUILD8* received the highest comprehension score (median value = 7), the users of *ODE2* obtained a slightly lower score (median = 6), while the users of *Chambers12* performed the worst (4.5). The Kruskal-Wallis test demonstrated that there was a statistically significant difference in the comprehension scores ($H = 42.264$, $df = 3$, $p\text{-value} < .0001$), and the post hoc test showed statistical differences between all the dictionary pairs except for *LDOCE*–*COBUILD* and *LDOCE*–*ODE*.⁷

The table also indicates the approaches to defining vocabulary and the values of receptive vocabulary load (RVL) of the definitions in each dictionary. The latter were calculated on the basis of the definitions to which the subjects were exposed in the comprehension task (and hence, some of the values differ from those reported in Chapter 4). From Table 16 and Figure 5, it transpires that the lower RVL of definitions, the higher comprehension score. This relationship is significant, with Kendall's $\tau = -.36$, p (one-tailed) < 0.0001 .

6. Spearman's $\rho = .15$

7. The critical difference ranged from 31.1 to 32.1 for *Chambers*–*LDOCE* and *COBUILD*–*ODE*, respectively, at the alpha level of $p = .05$. Most dictionary pairs exceeded this difference, which means that the differences between the comprehension scores for these dictionaries are statistically significant.

Table 16. Comprehension scores relative to the approach to defining vocabulary and RVL (total N = 215).

Definition source	N	Defining vocabulary	RVL (in word families)	Median comprehension score (scale 0–10)	Interquartile range
<i>LDOCE6</i>	55	controlled (with the use of an explicit set of DV items)	3,000	7	3
<i>COBUILD8</i>	50	uncontrolled (with no explicitly specified restrictions)	3,000	7	2
<i>ODE2</i>	54	uncontrolled (no restrictions whatsoever)	6,000	6	3
<i>Chambers12</i>	56	uncontrolled (no restrictions whatsoever)	18,000	4.5	3

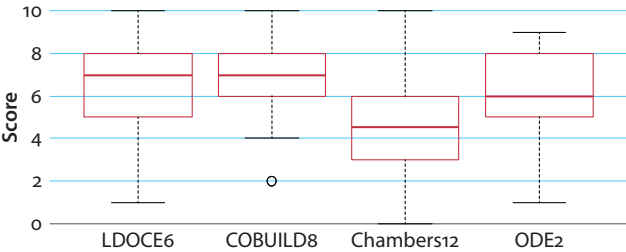


Figure 4. Comprehension scores by dictionary.

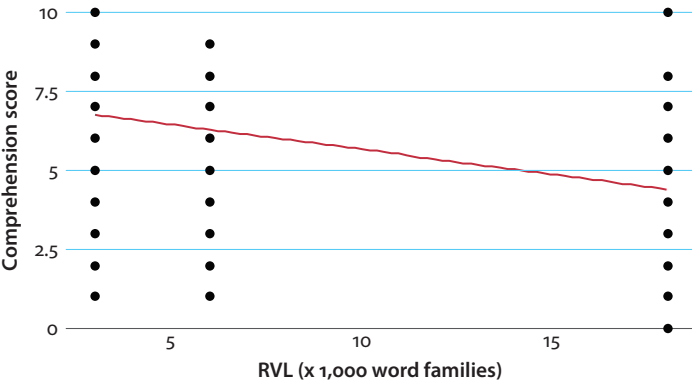


Figure 5. Correlation between comprehension score and receptive vocabulary load (RVL).

Another point of interest was to identify the definitions that were most difficult for the subjects to interpret. With this aim in view, the percentages of correct responses for each headword were computed. Results are displayed in Table 17. As can be seen, the number of correct responses varies depending on the dictionary to which the participants had access. For example, considering the first question, which concerns the meaning of *pungent*, one can see that 95% of the *LDOCE* users and only 59% of the *Chambers12* users provided correct responses. Counting the total percentage of correct responses for this word (83%), one can see that the question was relatively easy, because the word received the highest percentage of correct answers regardless of dictionary. On the other hand, *stingy* and *hobnob* were the most challenging (15% and 30% of total correct answers, respectively). Interestingly, as for *stingy*, low percentages of correct answers were observed for all the dictionaries, including *COBUILD8* and *LDOCE6* (10% and 25%, respectively). Some possible reasons will be discussed in Section 6.3.

Table 17. Percentages of correct responses.

Question No.	Headword	<i>LDOCE6</i>	<i>COBUILD8</i>	<i>Chambers12</i>	<i>ODE2</i>	Total
		subjects total: 55	subjects total: 50	subjects total: 56	subjects total: 54	subjects total: 215
1	<i>pungent</i>	95% (52)	90% (45)	59% (33)	91% (49)	83% (179)
2	<i>loophole</i>	87% (48)	88% (44)	55% (31)	72% (39)	75% (162)
3	<i>distinguished</i>	60% (33)	66% (33)	39% (22)	54% (29)	54% (117)
4	<i>hillbilly</i>	71% (39)	92% (46)	59% (33)	54% (29)	68% (147)
5	<i>lopsided</i>	78% (43)	82% (41)	57% (32)	67% (36)	71% (152)
6	<i>hobnob</i>	27% (15)	44% (22)	30% (17)	19% (10)	30% (64)
7	<i>stingy</i>	25% (14)	10% (5)	20% (11)	6% (3)	15% (33)
8	<i>obfuscate</i>	71% (39)	76% (38)	32% (18)	72% (39)	62% (134)
9	<i>loom</i>	85% (47)	90% (45)	46% (26)	74% (40)	73% (158)
10	<i>abject</i>	75% (41)	74% (37)	48% (27)	83% (45)	70% (150)

6.2.2 Comprehension of invented definitions

As in the previous test, the majority of the students did not know headwords 11–20 (Table 18). Also no correlation was found between the initial familiarity ratings and the comprehension scores (r close to 0).⁸ Therefore, it should come as

8. Spearman's rho=.03, p-value = 0.32

no surprise that there were subjects (59) who scored high on the comprehension test (8–10 points) in spite of low familiarity ratings (median = 1); and likewise, there were subjects (3) with low or middle comprehension score (0–6) but high familiarity score (4–5).

Table 18. Students’ familiarity with the words.

Median rating	1	2	3	4	5
Number of subjects	132 (61%)	50 (23%)	18 (8%)	11 (5%)	4 (2%)

The results of the comprehension task performed on invented definitions are presented in Table 19 and Figure 6. As can be seen in the table, the controlled and uncontrolled definitions present completely different levels of RVL: the former definitions need only 4 Base Lists to reach 98% coverage, whereas in the latter case 29 Base Lists are insufficient to reach that coverage. A Kruskal-Wallis test showed that the difference in comprehension score between the students groups was statistically highly significant ($H = 130.68$, $df = 1$, $p < .0001$)⁹. The high level of this significance indicates the importance of controlling a definition vocabulary for learners.

Table 19. The data in comprehension task with invented definitions (total N = 215).

Definitions	N	RVL (in word families)	Median score (scale 0–10)	Interquartile range
controlled	110	4,000	9	2
uncontrolled	105	>29,000	3	3

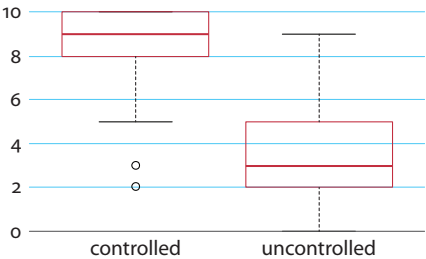


Figure 6. Comprehension of invented definitions.

9. $p\text{-value} < 2.2e-16$

Results of the analysis of correct responses are recorded in Table 20. As the table shows, all the headwords but *debauchery* received at least 81% of correct answers when defined in a controlled vocabulary, which is twice as much as when they were defined with no such restrictions. The exception will be presented in the discussion section.

Table 20. Percentages of correct responses.

Question No.	Headword	Definitions in controlled vocabulary	Definitions in uncontrolled vocabulary	Total
		subjects total: 110	subjects total: 105	subjects total: 215
11	<i>debauchery</i>	55% (61)	32% (34)	44% (95)
12	<i>redeem</i>	81% (89)	35% (37)	59% (126)
13	<i>invidious</i>	86% (95)	25% (26)	56% (121)
14	<i>impudent</i>	92% (101)	37% (39)	65% (140)
15	<i>solitude</i>	91% (100)	43% (45)	67% (145)
16	<i>stifle</i>	89% (98)	66% (69)	78% (167)
17	<i>infatuate</i>	89% (98)	29% (30)	60% (128)
18	<i>scrawny</i>	91% (100)	23% (24)	58% (124)
19	<i>peevish</i>	93% (102)	17% (18)	56% (120)
20	<i>superfluity</i>	95% (104)	32% (34)	64% (138)

6.2.3 Other factors affecting comprehension score

Apart from RVL and the approach to definition vocabulary, other factors likely to affect comprehension score were taken into consideration: the percentage of high-frequency and wide-range words, the percentage of synonym definitions, and average definition length. The results of correlation analyses are presented in Figures 7, 8, and 9. Table 21 presents the percentages of high-frequency and wide-range words in the definitions for headwords 1–10. The words were identified with the aid of AntWordProfiler as those covered by Base Lists 1–2. Figure 7 shows the relationship between comprehension score and the percentage of high-frequency and wide-range words. The correlation is statistically significant, $\tau=.36$, p (one-tailed) $<.0001$.

Table 21. Percentage of high-frequency and wide-range words, and comprehension of the definitions for headwords 1–10.

Dictionary	Percentage of high-frequency and wide-range words	Median comprehension score (scale 0–10)
<i>LDOCE6</i>	96.3%	7
<i>COBUILD8</i>	96.5%	7
<i>ODE2</i>	85.9%	6
<i>Chambers12</i>	74.1%	4.5

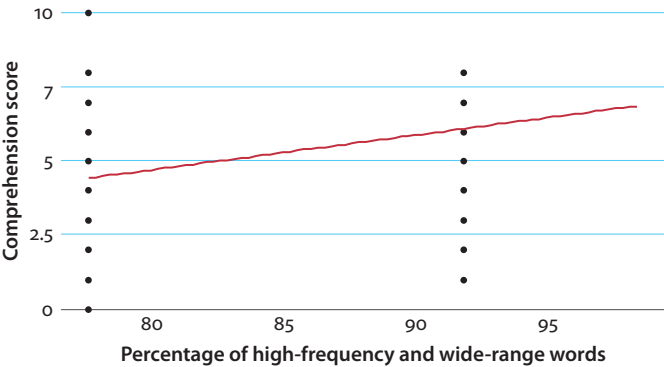


Figure 7. Correlation between comprehension score and the percentage of high-frequency and wide-range words in definitions.

Table 22 displays the data for the two other factors: percentage of synonym definitions and average definition length. As shown in Figures 8 and 9, the relationships are significant. There is a significant relationship between the percentage of synonym definitions and the comprehension score, $\tau = -.34$, p (one-tailed) < 0.0001 ; and definition length is significantly related to comprehension score, $\tau = .34$, p (one-tailed) < 0.0001 . In other words, comprehension of definitions increases with the lower number of synonym definitions and with longer definitions. The findings call into question the usefulness of synonym definitions in explaining meanings. At the same time, they emphasize the importance of analytical definitions for learners seeking to comprehend a word meaning.

Table 22. Percentage of synonym definitions, definition length, and comprehension of definitions for headwords 1–10.

Dictionary	Percentages of synonym definitions	Mean definition length (in words)	Median comprehension score (scale 0–10)
<i>LDOCE6</i>	10%	10.8	7
<i>COBUILD8</i>	0%	19.7	7
<i>ODE2</i>	20%	8.5	6
<i>Chambers12</i>	50%	5.8	4.5

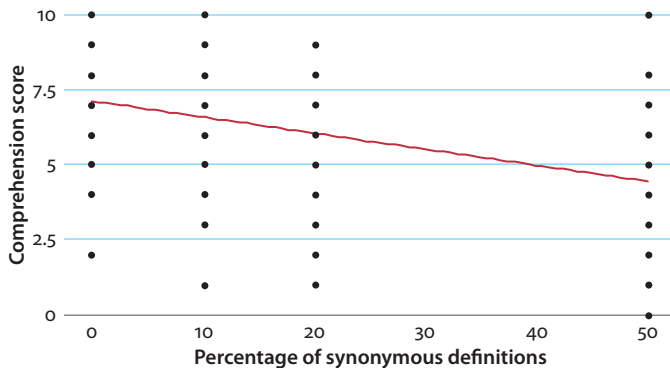


Figure 8. Correlation between comprehension score and the percentage of synonym definitions.

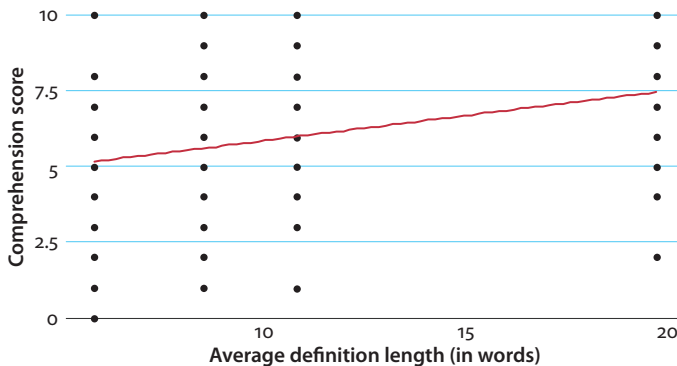


Figure 9. Correlation between comprehension score and average definition length.

6.2.4 Perception of definitions

The subjects evaluated a total of 196 definitions in terms of how helpful they were in conveying the meaning of the words, using the scale of 1 to 4. Table 23 shows subjects’ median ratings for the dictionaries. As can be seen in Table 23 and Figure 10, *Chambers12* received the lowest score (2.21) and *COBUILD8* the highest (3.36).

As the students’ measures were repeated for each dictionary, variation across the dictionaries in perceived helpfulness was computed using the Friedman’s ANOVA and a subsequent post hoc test. The Friedman’s ANOVA showed that there was a significant difference between the ratings across the dictionaries ($\chi^2 = 573.08$, $df = 3$, $p < 0.001$), and the post hoc test indicated that the differences were significant for all the dictionary pairs.¹⁰

Table 23. The data in perception task (N = 250).

Definitions	Median rating (scale 1–4)	Interquartile range
<i>LDOCE6</i>	3.16	0.48
<i>COBUILD8</i>	3.36	0.51
<i>Chambers12</i>	2.21	0.72
<i>ODE2</i>	2.71	0.52

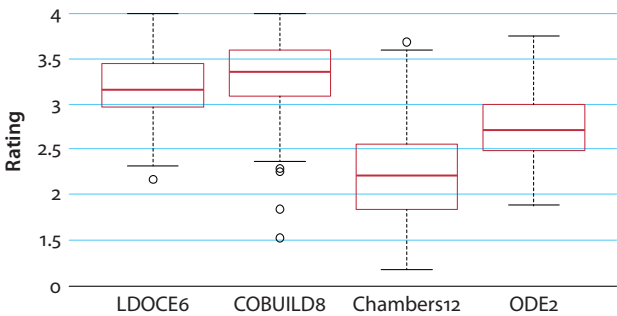


Figure 10. Students’ perception of definition helpfulness.

As demonstrated in Section 6.2.1, preference for a particular dictionary does not necessarily mean that its definitions are easier to understand than definitions in other dictionaries (see also Cumming et al. 1994; Nesi 1998).

10. The critical difference for all the pairs was 76.15 at $p = .05$.

6.2.5 Summary of the results

The findings of this experiment support all the hypotheses under consideration except for hypothesis H7:

1. the contribution of a controlled vocabulary to the learner's comprehension of definitions is statistically *highly significant* (H6);
2. contrary to hypothesis H7, the *COBUILD8* definitions are perceived as more helpful in explaining the meaning of words than the *LDOCE6* definitions;
3. the *ODE2* definitions are perceived as more helpful in explaining the meaning of words than the *Chambers12* definitions (H8);
4. definitions in *LDOCE6* and *COBUILD8* are equally comprehensible to foreign learners (H9);
5. definitions from *ODE2* are more comprehensible for foreign learners than definitions from *Chambers12* (H10);
6. the comprehension of definitions significantly correlates with RVL, the percentage of frequent and widely-used words in definitions, definition length, and the percentage of synonym definitions (H11).

6.3 Discussion

The definitions in *LDOCE6* and *COBUILD8* were equally successful in conveying the meanings of headwords. This finding can be accounted for by the identical values of RVL in these dictionaries (3,000 word families, respectively). The finding is consistent with the results of the study conducted by Nesi (1998) on earlier editions of these dictionaries (i.e. *LDOCE3* and *COBUILD2*). In other words, in the present study students' performance in the comprehension task was not affected by the different approach to defining vocabularies (i.e. with or without explicit restrictions). Nor was it influenced by the different definition format of the dictionaries (sentence definitions of *COBUILD8* vs. phrasal definitions of *LDOCE6*). This observation coincides with the results of Cumming et al.'s research on early editions of these dictionaries (i.e. *COBUILD1* and *LDOCE2*), which have comparable RVLs (5,000 and 6,000 word families, respectively; see 4.2). Finally, the comprehension scores were not affected by the subjects' initial familiarity with the headwords. The subjects who claimed familiarity were found to score low on comprehension. Similar cases of subjects claiming initial familiarity of words but scoring zero are reported by Nesi (1998).

As in Cumming et al.'s study (1994), the subjects rated higher the definitions from *COBUILD8* than those from *LDOCE6*. A possible reason for this

preference may have been the full-sentence format of the *COBUILD* definitions, which seems to have been more appealing to the learners than substitutional phrases of *LDOCE6*. Note, however, that strictly speaking both dictionaries use full-sentence definitions, but *COBUILD* does so consistently, while *LDOCE* occasionally. For example, all the *LDOCE6* definitions under study had a structure of a substitutable phrase, and only one had a full-sentence format. The qualities of full-sentence definitions can hardly be overlooked by dictionary users. On the other hand, considering the comparable levels of the receptive vocabulary loads of these dictionaries, it is doubtful that the selection of the DV had any influence on the learners' perceptions.

As regards the native-speaker dictionaries, they received significantly lower perception ratings than the learners' dictionaries, which is understandable given the different intended audiences. Noteworthy is the fact that the definitions in the native-speaker dictionaries differed in their perceived helpfulness, with the *ODE2* definitions being rated significantly higher than those in *Chambers12*.

In the comprehension task, the *ODE2* definitions were significantly more successful in conveying meaning than the *Chambers12* definitions. The fact that *ODE2* surpassed *Chambers12* in the comprehension task can be ascribed to the significantly lower RVL of the *ODE*'s definitions (6,000 vs. 18,000 word families, respectively), and the significantly higher percentage of high-frequency and wide-range words used in the definitions (85.9% in *ODE2* vs. 74.1% in *Chambers12*). These factors account for the success of the *ODE2* users, who performed in the comprehension task almost as well as the users of *LDOCE6* (6 and 7 points, respectively).

While it is evident that the differing success of *ODE2* and *Chambers12* in the comprehension task correlates with the choice of definition vocabulary, it also appears that other factors come into play, that is definition length and use of synonym definitions. Both of them were significantly correlated with comprehension score. The correlation was positive for definition length but negative for synonym definitions. This means that the former factor facilitates comprehension of definitions, while the latter inhibits it.

Native-speaker dictionaries vary in the provision of synonym definitions, which is noticeable when comparing *Chambers12* with *ODE2*. The high proportion of synonym definitions in *Chambers* (50%) is a key reason for its definitions being difficult to understand for the foreign learners (6.2.1). This proportion is more than twice as high as that of *ODE2* (20%).

Certainly, it is easier to control the choice of words for analytical definitions than for synonym definitions. While the former definition is a paraphrase of a word being defined in terms of the words that are semantically simpler and usually more frequent than the word defined, a synonym definition is based on a different

assumption, namely sameness or similarity of meaning. Finding a near-synonym is difficult when the choice of defining words is restricted to a subset of the lexicon. In selecting words for a DV, lexicographers avoid inclusion of a synonym for a general word already on the list, but even if such an extra word is permissible it is usually less frequent and more restricted in range than the other word (e.g., *start* vs. *commence* or *initiate*).

Although synonym definitions take up little space, they may lead to the learner's misconceptions regarding the meaning of the word being defined (Jain 1981; Nesi and Maera 1994). Defining by synonyms does not offer the same opportunity for comprehending definitions as the classical definition. The latter offers substitution in the form of a superordinate term (Fischer 1991), which is normally more common and easier to understand than the word being defined. There is no doubt that although definition length correlates with comprehension score, the correlation is not unconditional, as extremely long and convoluted definitions do not warrant that the correlation will still hold. All the same, definition length and format are important criteria of dictionary design, strictly related not only to accessibility of definitions, but also the envisaged audience, situations of dictionary use (production vs. reception), types of texts for decoding, etc.

Examining students' responses in the comprehension task, one notices that the questions about the meaning of *stingy a.* and *hobnob v.* were the most challenging. Analysis of incorrect answers regarding the latter word shows that the majority of the subjects who committed this error indicated option "d", that is, "wstąpić do towarzystwa, klubu lub partii" ("join a society, club or party"), rather than indicating "c", "kręcić się koło kogoś, przestawać z kimś" ("to associate with, to keep company"). The incorrect response was given especially by the users of *ODE2* (57%), *LDOCE6* (46%), and *COBUILD8* (36%), who must have overinterpreted their definitions, especially the parts that specified the people one typically hobnobs with (i.e. "of perceived higher social status", "rich and powerful", "in a higher social position"), assuming that joining a society, club or party opens up an opportunity for meeting important people. A lower percentage of this answer was found in the *Chambers12* users (26%), which may be attributed to the fact that their definition, in contrast to the others, does not mention a high status of people. It is worth noting that in the case of the *ODE2* students, their misinterpretation of the main part of the *ODE2* definition ("mix socially" in "mix socially, especially with those of perceived higher social status") may have given rise to another error: "wmieszać się w tłum" ("mix with a crowd") in the option "b". This error was committed by 26% of the *ODE2* users. Here the students not only failed to recognize the non-central metaphorical meaning of "mix socially", but also ignored the remaining part of the definition that specified the people one typically hobnobs with ("those of perceived higher social status"). This shows that use of a defining

word in a non-central meaning may cause confusion and that careless reading of definitions leads to errors.

As for the other problematic headword, *stingy* (*with words*), there were predominantly two types of errors. One was answer “c”: “pazerny na pieniądze” (“mean with money”), and the other, answer “a”: “używający obraźliwych słów” (“using offensive words”). The former error, which was common among the users of *LDOCE6* (75% of answers) and *COBUILD8* (85%), arose from the fact that they failed to notice that the question regarded the meaning of *stingy with words*, not *money*. This again must have been caused by students’ cursory reading of the question, and inability to adjust the dictionary information to the task at hand. The latter error (answer “a”) was found mainly in *ODE2* (57% of answers), and was presumably due to students’ misinterpretation of the synonym definition: “mean; ungenerous”. It has already been demonstrated earlier that learners tend to struggle with such definitions. The case of *stingy* illustrates that students’ success in comprehension tasks depends on their reading skills and definition format.

The value of a controlled vocabulary for the learner cannot be overestimated. This is clear from the experiment with invented definitions, which showed a statistically highly significant difference in comprehension score between the readers of controlled and of non-controlled definitions. However, the experiment showed that simplification of definitions may lead to their misinterpretation by learners. Questions 11–20 were generally easy for the learners who had access to controlled definitions, but one word with relatively fewer correct responses (55%) was *debauchery*. Interestingly, incorrect responses for this word were dominated by choice “a”: “bezwzględny egoizm” (“ruthless egoism”) (34%). This may have been due to the simplification of the definition, “the habit of allowing oneself to behave immorally”, which, being less precise and more general than the non-simplified one (“the habit of indulging in dissolute practices”), allowed for two responses “a” and “d”. Note that both responses represent a type of *immoral behaviour*. This case demonstrates that precision of a controlled definition may sometimes be sacrificed for its greater comprehensibility (cf. Stein 1979, Fox 1989). Nevertheless, this seems to be only a minor downside which is far outweighed by advantages.

6.4 Conclusions

As summarized in 6.2.5, the findings reported in this chapter support all the hypotheses except for hypothesis H7. One of the main findings that emerge from the analysis of invented definitions is that the learners who have access to controlled definitions are *highly* significantly more successful in their comprehension than the learners who use the definitions written with no lexical restrictions

whatsoever. However, it seems that the approach to the selection of definition vocabulary (i.e. with or without explicit restrictions) is less relevant to the learner's success than the lexical burden of definitions (RVL). Although the learners perceived the *COBUILD8* definitions as more helpful than the *LDOCE6* definitions, there was no evidence for superiority of one dictionary over the other in conveying the meanings of words. This finding can be attributed to the identical RVL of the definitions to which the users of these dictionaries were exposed (3,000 word families).

As regards native-speaker dictionaries, the study showed considerable differences between *Chambers12* and *ODE2* with respect to both comprehension and perception of their definitions by foreign learners. From a learner's perspective, the latter dictionary has the advantage over the former of using fewer synonym definitions and more common defining words. This comparison shows that native-speaker dictionaries do not offer the same opportunities for learners to understand a difficult word, even when they are used with no supporting context. Such factors as frequent and widely-used defining words, definition type, and definition length affect comprehension of definitions. Thus, it seems that advanced learners will be better off using *ODE2* than *Chambers12* if they do not have an EFL dictionary at hand.

What is more, the fact that the *ODE2* users performed almost as well as the *LDOCE6* users on multiple choice questions with no contextual aid runs counter to the seemingly reasonable prediction that native-speaker dictionaries provide little help for learners in such tasks (cf. McCreary and Dolezal 1999).

General conclusions

1. Main findings

This monograph focused on the under-researched issue of controlling the vocabulary of definitions in English dictionaries. The aim of the book was threefold: firstly, to investigate the role of vocabulary control in how learners understand definitions, secondly, to examine relevant lexicographic practices, and thirdly, to explore their implications for definition writing in learner's dictionaries. I conducted a series of studies of lexical content of dictionary definitions and the studies of comprehensibility and perception of definitions by learners of English. The results are presented in Chapters 4 to 6.

A central concern of Chapter 4 was to compare definitions from various dictionaries with regard to receptive vocabulary load (RVL) in order to estimate the likelihood of their vocabularies being known by learners. The study demonstrated, unsurprisingly, that the learners' dictionaries show incomparably lower RVLs than the native-speaker dictionaries (3,000 and 26,000 word families, respectively). This means that comprehension of definitions in the former dictionaries requires significantly less knowledge of definition vocabulary than is the case in the latter dictionaries. This comparison also shows that learner's dictionaries go a great deal further than native-speaker dictionaries to achieve comprehensibility.

Chapter 6 provided further evidence in support of the fact that vocabulary control plays a key role in the learner's understanding of definitions. The comprehension test performed on 215 learners using invented definitions demonstrated that a controlled vocabulary contributes highly significantly to the learner's comprehension of definitions. This suggests that a consistent vocabulary control—in the broad sense of the word—is not merely useful but indispensable for the learner's successful understanding of definitions.

The other studies presented in Chapters 4 to 6 examine lexicographic policies and practices in the field of vocabulary control. The study of RVL in Chapter 4 showed that although the vocabulary of definitions in learner's dictionaries presented comparable levels of difficulty for learners (3,000 and 4,000 word families), the lower threshold of 3,000 word families was reached only by the dictionaries using a restricted defining vocabulary (*LDOCE6* and *CALD4*). This finding tips the balance in favor of West's model of vocabulary control, which imposes greater discipline on lexicographers in the selection of words for definitions.

This finding implies that West's model of restricting the vocabulary to a limited set of well-chosen items is more likely to make definitions easier to understand for learners than Hornby's model of controlling this vocabulary by merely using words that are simpler than the word being defined. However, the claim was not validated by the user-oriented study of comprehension of 10 definitions from *LDOCE6* and *COBUILD8* (Chapter 6), which demonstrated that the definitions were equally successful in conveying meanings to the learners in spite of the different approaches to vocabulary control (with and without explicit restrictions, respectively). Nevertheless, the success of the learners with these dictionaries should be accounted for by the identical RVL of the definitions tested (3,000 word families). The user-research demonstrated that both approaches to vocabulary control can ensure successful comprehension of definitions.

The success of the two approaches lies in the systematic execution of vocabulary restrictions, whether imposed prior to writing definitions (by means of an explicit list of allowable items) or applied during the writing process itself. The study presented in Chapter 5 shows that recent dictionaries are more effective in this regard than earlier ones, irrespective of the approach to vocabulary control. Improvements are noticeable in the limited use of technical terms, stylistically marked words, and other potentially difficult words and expressions. In contemporary lexicography restrictions on the form and, in a much less degree, the meaning of a word can be applied with the aid of specialist software, statistical tools, and a corpus evidence. Editors can easily monitor parameters of the words selected for definitions. While computer software does not prevent lexicographers from using words in idiomatic combinations or non-central meanings, computational techniques can greatly facilitate the verification process. This process was more difficult a few decades ago, when editors did not have specialist software at their disposal. Given the fact that practically until the early 1970s dictionaries were compiled with no or little support of computers (see 3.3.2.2), the task of writing definitions consistently with words more basic than the word being defined was daunting. The lack of automatic means to monitor definition vocabularies partly explains the fact that the definitions of *ISED* and the other early editions of *OALD* are lexically more challenging for learners than definitions in the other EFL dictionaries.

The user-oriented study presented in Chapter 6 pointed to other factors, besides RVL and the approach to vocabulary control, that may affect comprehension of definitions. They include the content of the most frequent and widely used words in definitions (e.g., those covered by Base Lists 1 and 2), definition length, and definition type. The effect of these factors was noticeable in the comprehension task taken by learners using native-speaker dictionaries. The study showed that the learners using the *ODE2* definitions performed significantly better than the learners who had access to the *Chambers12* definitions. Among the reasons

is the fact that the former dictionary uses fewer synonym definitions and more high-frequency and wide-range words in its definitions. Apparently defining by synonym does not make it feasible for a lexicographer to control the vocabulary of a definition and for a learner to understand fully the meaning of the word being defined. The fact that the *ODE2* users performed nearly as well as the *LDOCE6* users in this task suggests that the former dictionary caters relatively well to the needs of less proficient users even though it is designed with native speakers in mind. This finding may result from the fact that the publisher aims to reach a large audience in the world dominated by the English language.

One of the studies presented in Chapter 4 revealed that there is no specific DV that will be sufficient to define within reasonable limits and in a precise way all the lexical units in the language. Firstly, definitions of numerous domain-specific words (and the words crossing domain boundaries) such as *gherkin*, *grouse*, and *lava*, require the use of other similar words that do not qualify as members of a DV list because they are neither sufficiently frequent nor widespread in the general language. Secondly, the dictionaries do not cover evenly the same specialist domains. Furthermore, in the interest of learning economy, the dictionaries exclude from their lists semantically predictable derivatives, compounds, and combinations of words, and they do so in a non-uniform way. All these factors contribute to the DV lists being varied in lexical content.

Yet, in spite of considerable differences between the dictionaries, it was found that they agree on the 1,520, which I called the Core Defining Vocabulary (CDV). The CDV consists of predominantly high-frequency and wide-range words, stylistically unmarked, with a high potential for substitution for other words. The CDV can be considered as most useful for definers in writing definitions and for learners to understand them. This lexical core has a long-lasting value for definitions, as it is based on at least 30 years of lexicographers' experience in preparing DVs. The fact that 92% of this vocabulary overlaps with the GSL (1953) shows that it is not only resistant to change over time but also necessary for general teaching and learning purposes. The core vocabulary is small, which means that its acquiring can be a realistic learning goal for less proficient students who want to attain the advanced level.

2. Implications for definition writing and DV design

The analyses presented in this book lead to practical conclusions. Below is a selection of guidelines for designers of a defining vocabulary. This is not a complete set but a selection of optimal solutions which emerged in the course of examining definitions. They apply to general-purpose dictionaries for advanced learners.

1. Because it is impossible to predict in advance the entire set of lexical units necessary for all definitions in a general-purpose dictionary for advanced learners, it would be misleading to claim that 2,000 or 3,000 words are sufficient for that purpose. However, it is absolutely critical for learners to know the core DV, and pedagogically desirable to teach this vocabulary to learners of English at a low level of proficiency. The core DV, which is estimated to be around 1,500 words (4.3.4), ensures that such learners get the best return for their learning effort. The unpredictability of the entire set of defining items does not diminish the usefulness of the core DV for learners.
2. Both approaches to vocabulary control (with and without an explicit set of DV items) are capable of reducing RVL of definitions to a level sufficiently low for advanced learners. However, the fact that only the dictionaries with a restricted defining vocabulary achieved the lowest level (3,000 word families) indicates that these dictionaries are more suited to less competent learners. For advanced learners, the choice between these approaches appears to be less relevant than the systematic execution of the principle of defining with the words likely to be known by such users. The words can be identified as belonging to Nation's Base Lists 1 to 4, because the lists cover 98% of the defining vocabulary in contemporary learners' dictionaries. This coverage is sufficient for advanced learners to ensure adequate comprehension of definitions.
3. Ideally, a DV ought to consist of only those lexical units that the target user is expected to know. However, in the interest of accuracy they need to be supplemented with a certain amount of less common words, especially those which are necessary in definitions of specialist and domain-specific terms. Without them definitions will be lengthy or imprecise. Some of these words may be indispensable only in a handful of definitions. In paper dictionaries they need to be explained through glosses, and in electronic dictionaries by hyperlinks. Nevertheless, for ease of reference, the use of rare and specialist terms in definitions should be kept under control even in electronic dictionaries.
4. Definitions ought to be written using words in central or core meanings. This is beyond dispute, but which meanings are central is a matter of common sense rather than of the editor's arbitrary decision. Prescribing meanings in the form of an explicit list of allowable words and their senses does not warrant that the words will be used in these senses consistently throughout a dictionary. This is because word sense disambiguation is an intractable problem (see 2.4). While off-the-list items can easily be detected by computer software, it is more difficult to determine automatically which items are used in definitions in peripheral or rare senses. Considering the fact that high-frequency

words tend to have rich polysemy, the use of the words in more than one meaning is unavoidable. Thus, it is perhaps more practical to equip lexicographers with corpus-based statistics on the distribution of word meanings and let them make decisions as to which meanings are central.

5. In the interest of the naturalness of the defining language, common phrasal verbs ought to be included in a controlled defining vocabulary. Although they are non-compositional units, they are frequent in general English and should not be a problem for an advanced learner. For the same reason, it is justifiable to include in the DV list more than one lexeme for the same concept as long as the lexeme is known by a learner. For example, along with the semantically general word *false*, which has a great potential for substitution (see 2.4), it is worth using *artificial* because it naturally collocates with *intelligence*, *light*, *substances*, *materials*, *limbs*, *insemination*, *ventilation*, *pitch*, and *respiration*. The use of typical collocations contributes to definitions being natural.
6. When compiling a defining vocabulary, lexicographers should assume that a large part of the learner's receptive vocabulary conforms to the principles of word formation and meaning extension. This is the passive vocabulary that is easy for the learner to recognize and understand, but may be difficult to recall in production. In order to reduce a learning effort, there is no need to extend the DV list by including semantically transparent derivatives and compounds as long as their bases and compound elements are on the list. By the same token, there is no need to include meanings that are easily derivable from primary ones. This is the case of regular polysemy; for example, it is perfectly justified to use *can*, *bottle* and *cup* in the sense of "liquid" along with the sense of "container".
7. Lexicographers ought to avoid defining solely by synonym, and combining synonyms with analytical definitions, unless the former are marked explicitly as separate components of an entry.
8. The use of parenthetical glosses to explain the meanings of non-DV words is an example of best practice in learner's lexicography. They are crucial for learners, and even non-expert native speakers, in the comprehension of many definitions of specialist and natural-kind terms. However, glossing may not always be sufficient for a learner to understand a definition or to identify the object being described. Whenever this strategy fails to fulfill those aims, it is worth considering the incorporation of elements of the bilingual dictionary, such as L1 equivalents for the word being defined or for a rare word used in definitions. This is a promising line of development for electronic dictionaries.

3. Further research

Further research is needed on controversial aspects of a DV. An interesting research avenue is a comparison of definitions of different groups of words, including those that are difficult to explain within a DV (e.g., abstract and concrete nouns, natural-kind terms). It would be interesting to see to what extent a defining vocabulary improves or hampers their comprehensibility. Such a comparison should be extended to include bilingual and bilingualized dictionaries. Also it would be worthwhile to perform a user-oriented study to explore whether the syntax of definitions written with a DV is more challenging for learners as compared to that of unrestricted definitions.

APPENDIX 1

The Core Defining Vocabulary (CDV)

a	agreement	appear	backwards
ability	aim	appearance	bad
about	air	apple	bag
above	aircraft	approval	bake
accept	airport	area	balance
acceptable	alcohol	argue	ball
accident	alive	argument	bank
according to	all	arm	bar
account	allow	army	base
across	almost	around	bath
act	alone	arrange	beak
action	along	arrangement	bear
active	alphabet	arrive	beat
activity	already	art	beautiful
actor	also	article	beauty
add	although	artificial	because
addition	always	as	become
address	among	ashamed	bed
admire	amount	ask	beer
admit	ancient	at	before
adult	and	atom	begin
advantage	anger	attack	beginning
advertise	angle	attempt	behave
advertisement	angry	attention	behind
advice	animal	attract	belief
advise	annoy	attractive	believe
after	another	autumn	bell
afternoon	answer	average	below
again	any	avoid	belt
against	anyone	awake	bend
age	anything	away	between
ago	anywhere	baby	bicycle
agree	apart	back	big

bird	bring	ceremony	coffee
birth	broadcast	certain	coin
bite	brother	chain	cold
bitter	brown	chair	collect
black	brush	chance	college
blade	build	change	combination
blame	building	character	combine
blind	bullet	charge	come
block	burn	chase	comfortable
blood	bury	cheap	common
blow	bus	cheat	company
blue	bush	cheese	compare
board	business	chemical	comparison
boat	busy	chemistry	compete
body	but	chest	competition
boil	butter	chicken	complain
bomb	button	child	complaint
bone	buy	chocolate	complete
book	by	choice	computer
boot	cake	choose	condition
border	calculate	church	confidence
born	call	cigarette	confident
borrow	calm	cinema	connect
both	camera	circle	conscious
bottle	can	city	consider
bottom	capital	claim	contain
bowl	car	class	continue
box	card	clean	continuous
boy	care	clear	control
brain	careful	clever	conversation
branch	careless	climb	cook
brave	carry	clock	copy
bread	case	close	corner
break	castle	cloth	correct
breast	cat	clothes	cost
breath	catch	clothing	cotton
breathe	cause	cloud	cough
brick	cell	coal	could
bridge	central	coast	count
bright	century	coat	country

course	defend	dress	engine
court	degree	drink	enjoy
cover	delay	drive	enjoyment
cow	delicate	drop	enough
cream	department	drug	enter
creature	depth	drum	entertain
crime	describe	drunk	entertainment
criminal	description	dry	entrance
crop	desert	during	envelope
cross	deserve	dust	equal
crowd	destroy	duty	escape
cruel	destruction	each	especially
crush	detail	ear	establish
cry	determination	early	even
cup	determined	earn	evening
cupboard	develop	earth	event
cure	dictionary	east	ever
current	die	eastern	every
curtain	difference	easy	everyone
curve	different	eat	everything
custom	difficult	edge	everywhere
customer	difficulty	educate	evil
cut	dig	education	exact
damage	direct	effect	examination
dance	direction	effective	examine
danger	dirt	effort	example
dangerous	discover	egg	excellent
dark	discovery	eight	except
date	dish	either	exchange
daughter	distance	elect	exciting
day	divide	election	excuse
dead	do	electric	exercise
death	doctor	electricity	exist
decay	dog	else	existence
decide	door	employ	expect
decision	doubt	employer	expensive
decorate	down	empty	experience
decoration	draw	encourage	explain
deep	drawer	end	explanation
defeat	dream	enemy	explode

explosion	finish	freeze	great
express	fire	fresh	green
expression	firm	friend	ground
extreme	first	friendly	group
eye	fish	frighten	grow
face	fit	frightening	growth
fact	five	from	guard
factory	fix	front	guess
failure	flag	fruit	guest
fair	flame	full	gun
fall	flash	funeral	habit
false	flat	fur	hair
familiar	flesh	furniture	half
family	flight	future	hammer
famous	float	game	hand
far	flood	garden	handle
farm	floor	gas	hang
farmer	flour	gate	happen
fashion	flow	general	happy
fashionable	flower	generous	hard
fast	fly	gentle	harm
fasten	fold	get	harmful
fat	follow	girl	hat
father	food	give	hate
fault	foot	glass	have
fear	football	glue	he
feather	for	go	head
feed	force	god	health
feel	foreign	gold	healthy
feeling	forest	good	hear
female	forget	goods	heart
fence	forgive	govern	heat
fever	fork	government	heaven
few	form	gradual	heavy
field	formal	grain	height
fight	forward	grammar	help
fill	four	grandfather	helpful
film	frame	grandmother	her
find	free	grass	here
finger	freedom	grateful	herself

hide	improve	jump	level
high	improvement	just	lid
hill	include	keep	lie
him	income	key	life
himself	increase	kick	lift
his	industrial	kill	light
historical	industry	kilogram	like
history	infect	kind	likely
hit	infection	king	limit
hold	infectious	kiss	line
holiday	influence	kitchen	lip
holy	information	knee	liquid
home	ink	knife	list
honest	insect	knock	listen
hope	inside	knot	literature
horn	instead	know	little
horse	instruction	knowledge	live
hospital	instrument	lack	local
hot	intend	lake	lock
hotel	intention	land	long
hour	interest	language	look
house	interesting	large	loose
how	international	last	lose
human	interrupt	late	loss
humorous	into	laugh	lot
hungry	introduce	law	loud
hunt	introduction	lawyer	love
hurry	invent	leaf	low
hurt	invitation	lean	loyal
husband	invite	learn	luck
ice	iron	least	lung
idea	island	leather	machine
if	it	leave	magazine
ill	its	left	magic
image	job	leg	main
imaginary	join	legal	make
imagination	joke	lend	male
imagine	journey	length	man
importance	judge	less	manage
important	juice	letter	manager

manner	moment	net	one
many	money	never	onion
map	month	new	only
march	moon	news	open
mark	moral	newspaper	operate
market	more	next	operation
marriage	morning	night	opinion
marry	most	nine	oppose
mass	mother	no	opposite
match	motor	noise	opposition
material	mountain	none	or
matter	mouse	nonsense	orange
me	mouth	nor	order
meal	move	north	to
mean	much	northern	that
meaning	multiply	nose	ordinary
measure	murder	not	organ
meat	muscle	nothing	organization
medical	music	notice	origin
medicine	musician	now	other
meet	must	nowhere	our
meeting	my	number	ours
member	mysterious	nurse	out
memory	mystery	nut	outer
mention	nail	obey	outside
message	name	object	over
metal	narrow	obtain	owe
method	nation	occasion	own
middle	national	of	page
military	nature	off	pain
milk	navy	offensive	painful
mind	near	offer	paint
minute	nearly	office	painting
mirror	necessary	officer	pair
miss	neck	official	pale
mistake	need	often	pan
mix	needle	oil	paper
mixture	neither	old	parallel
model	nerve	on	parent
modern	nervous	once	parliament

part	plan	preparation	quality
particular	plant	prepare	quarter
partner	plastic	present	queen
party	plate	preserve	question
pass	play	press	quick
passage	pleasant	pressure	quiet
passenger	please	pretend	race
past	pleased	prevent	radio
path	pleasure	price	railway
patient	pocket	priest	rain
pattern	poem	principle	raise
pause	poetry	print	range
pay	point	prison	rank
payment	poison	prisoner	rare
peace	poisonous	private	rate
peaceful	pole	prize	rather
pen	police	problem	reach
pencil	polite	process	read
people	politics	produce	ready
perfect	political	product	real
perform	politician	production	really
period	pool	profit	reason
permission	poor	promise	reasonable
person	popular	pronounce	receive
personal	port	pronunciation	recent
persuade	position	proof	recently
pet	possession	property	recognize
photograph	possibility	protect	record
phrase	possible	protection	red
physical	possibly	proud	reduce
piano	post	provide	reduction
picture	potato	public	refuse
piece	pour	pull	regular
pig	powder	punish	relation
pile	power	punishment	religion
pin	powerful	pure	religious
pink	practical	purple	remark
pipe	praise	purpose	remember
place	pray	push	remove
plain	prayer	put	rent

repair	salt	sheet	sleep
repeat	same	shelf	slide
report	sand	shell	slight
represent	satisfaction	shelter	slope
request	save	shine	slow
respect	say	ship	small
responsible	school	shirt	smell
rest	science	shock	smile
restaurant	scientific	shoe	smoke
result	scientist	shoot	smooth
return	sea	shop	snake
reward	search	short	snow
rice	season	should	so
rich	seat	shoulder	soap
ride	second	shout	social
right	secret	show	society
ring	see	side	sock
rise	seed	sideways	soft
risk	seem	sight	soldier
river	sell	sign	solid
road	send	signal	some
rock	sentence	silent	something
roll	separate	silly	sometimes
roof	serious	silver	somewhere
room	servant	similar	son
root	serve	simple	song
rope	service	since	soon
rough	set	sincere	sore
round	seven	sing	sorry
row	several	single	sound
royal	severe	sink	soup
rub	sew	sister	sour
rubber	sex	sit	south
rude	sexual	situation	southern
rule	shake	six	space
sad	shape	size	speak
safe	share	skill	special
safety	sharp	skin	speech
sail	she	skirt	speed
sale	sheep	sky	spell

spend	student	television	too
spin	study	tell	tool
spoil	style	temperature	tooth
spoon	subject	tent	top
sport	substance	test	total
spot	succeed	than	touch
spring	success	the	towards
square	successful	their	tower
stage	such	them	town
stair	suck	themselves	toy
stamp	sudden	then	track
stand	suffer	there	trade
standard	sugar	therefore	traffic
star	suggest	they	train
start	suit	thick	transparent
state	suitable	thief	travel
station	summer	thin	treat
stay	sun	thing	treatment
steal	supply	think	tree
steam	support	third	trick
steep	surface	thirsty	tropical
step	surprise	this	trouble
stick	surround	thought	trousers
sticky	swallow	threat	true
stiff	sweet	threaten	trust
still	swell	three	truth
stomach	swim	throat	try
stone	sympathy	through	tube
stop	system	throw	tune
store	table	thumb	turn
storm	tail	ticket	twice
story	take	tidy	twist
straight	talk	tie	type
strange	tall	tight	typical
stream	taste	time	tyre
strength	tax	title	ugly
stretch	tea	today	under
string	teach	toe	understand
strong	team	together	uniform
structure	tear	tongue	unit

unite	wait	wet	with
universe	wake	what	within
university	walk	whatever	without
until	wall	wheel	woman
up	want	when	wood
upper	war	where	wooden
upset	warm	which	wool
urgent	warmth	while	word
us	warn	white	work
use	wash	who	world
useful	waste	whole	worry
usual	watch	why	worse
valley	water	wide	worth
valuable	wave	width	would
value	way	wife	wrap
various	we	wild	wrist
vegetable	weak	will	write
vehicle	weapon	willing	wrong
very	wear	win	year
view	weather	wind	yellow
village	week	window	yes
violence	weigh	wine	yet
violent	weight	wing	you
visit	welcome	winter	young
voice	well	wire	your
vote	west	wise	yours
waist	western	wish	zero

APPENDIX 2

Word Familiarity Page

Celem badania jest ocena słowników a nie studentów, dlatego ankieta jest anonimowa.¹

Zaznacz na skali, w jakim stopniu znasz te słowa:

1 = w ogóle nie znam tego słowa

5 = znam słowo bardzo dobrze²

<i>pungent</i>	1	2	3	4	5
<i>loophole</i>	1	2	3	4	5
<i>distinguished</i>	1	2	3	4	5
<i>hillbilly</i>	1	2	3	4	5
<i>lopsided</i>	1	2	3	4	5
<i>hobnob</i>	1	2	3	4	5
<i>stingy</i>	1	2	3	4	5
<i>obfuscate</i>	1	2	3	4	5
<i>loom</i>	1	2	3	4	5
<i>abject</i>	1	2	3	4	5
<i>debauchery</i>	1	2	3	4	5
<i>redeem</i>	1	2	3	4	5
<i>invidious</i>	1	2	3	4	5
<i>impudent</i>	1	2	3	4	5
<i>solitude</i>	1	2	3	4	5
<i>stifle</i>	1	2	3	4	5
<i>infatuate</i>	1	2	3	4	5
<i>scrawny</i>	1	2	3	4	5
<i>peevish</i>	1	2	3	4	5
<i>superfluity</i>	1	2	3	4	5

1. [The aim of this study is to evaluate dictionaries, not students, so the questionnaire is anonymous.]

2. [Indicate how much you are familiar with each word. Use the scale from 1 to 5. 1 = "I don't know the word" 5 = "I know the word very well"]

Task 1³, version 1 (with *LDOCE6* definitions)

Zadanie 1. Przeczytaj definicję wytłuszczonego wyrazu w dolnej części ramki i zakresl jedną odpowiedź (a, b, c, d), która najbardziej odpowiada znaczeniu tego wyrazu. Jeśli nie wiesz, zaznacz odpowiedź e.⁴

1. **Pungent** smell to:

- (a) przyjemny zapach
- (b) mdły zapach
- (c) drażniący zapach
- (d) ledwo wyczuwalny zapach
- (e) nie wiem

pungent having a strong taste or smell

2. **Loophole** to:

- (a) niechęć pracodawcy do sporządzania umów
- (b) brak przepisów prawnych normujących jakąś sytuację lub błąd w przepisach
- (c) natychmiastowe rozwiązanie umowy z pracownikiem
- (d) środki ostrożności podejmowane przez pracodawców
- (e) nie wiem

loophole a small mistake in a law that makes it possible to avoid doing something that the law is supposed to make you do

3. **Distinguished** oznacza:

- (a) poważany
- (b) ilustrowany
- (c) apodyktyczny
- (d) mało charakterystyczny
- (e) nie wiem

distinguished successful, respected, and admired

3. In the multiple choice test, answers “a” and “b” under **debauchery** are translations of McCreary and Amacker’s test (2006).

4. [Task 1. Read the definition of the word in bold, which is given at the bottom of each box, and circle one of the answers (a, b, c, d) which bests reflects the meaning of the word. If you don’t know, indicate answer e.]

4. Hillbilly to:

- (a) rustykalny urok
- (b) prostak
- (c) osoba nieodgadniona
- (d) osoba skromna
- (e) nie wiem.

hillbilly an insulting word meaning an uneducated poor person who lives in the mountains

5. Lopsided to:

- (a) wąski
- (b) znajdujący się przy samym brzegu czegoś
- (c) odwrócony do góry nogami
- (d) pochylony do boku
- (e) nie wiem

lopsided having one side that is lower or heavier than the other

6. Hobnob oznacza:

- (a) przebywać z rodziną
- (b) wmieszać się w tłum
- (c) kręcić się koło kogoś, przestawać z kimś
- (d) wstąpić do towarzystwa, klubu lub partii
- (e) nie wiem

hobnob to spend time talking to people who are in a higher social position than you

7. Stingy with words oznacza:

- (a) używający obraźliwych słów
- (b) wrażliwy na czyjeś słowa
- (c) pazerny na pieniądze
- (d) mało mówny
- (e) nie wiem.

stingy not generous, especially with money

8. Obfuscate to:

- (a) połączyć
- (b) rozwścieczyć
- (c) zaciekawić
- (d) zaciemnić
- (e) nie wiem

obfuscate to deliberately make something unclear or difficult to understand

9. Loom to:

- (a) szybko następować po sobie, galopować
- (b) zbliżać się wielkimi krokami
- (c) odwołać
- (d) naświetlić
- (e) nie wiem

loom if a problem or difficulty looms, it is likely to happen very soon

10. Abject failure to:

- (a) przypadkowe niepowodzenie
- (b) żalosne, okropne niepowodzenie
- (c) nieuchronne niepowodzenie
- (d) błahe, drobne niepowodzenie
- (e) nie wiem

abject the state of being extremely poor, unhappy, unsuccessful etc

11. Debauchery to:

- (a) bezwzględny egoizm
- (b) namiętne pragnienie
- (c) chciwość
- (d) rozpasanie, rozpusta
- (e) nie wiem

debauchery the habit of indulging in dissolute practices

12. To redeem a mistake oznacza:

- (a) uniknąć błędu
- (b) naprawić błąd
- (c) skrytykować błąd
- (d) doszukiwać się błędu
- (e) nie wiem

redeem to atone (a mistake)

13. An invidious topic to temat:

- (a) który może być nieprzyjemny
- (b) który dotyczy choroby
- (c) który narzuca komuś sposób myślenia
- (d) który trudno jest określić
- (e) nie wiem

invidious likely to entail ill-will

14. *An impudent person* to osoba:

- (a) która stroni od innych ludzi
- (b) która jest poważna
- (c) która jest wyjątkowo uparta
- (d) która jest bezczelna i zuchwała
- (e) nie wiem

impudent showing unblushing insolence to other people

15. *Solitude* to:

- (a) poczucie winy
- (b) solidarność
- (c) odosobnienie
- (d) poczucie bezpieczeństwa
- (e) nie wiem

solitude seclusion

16. *To stifle a smile* oznacza:

- (a) uśmiechnąć się szeroko
- (b) wybuchnąć śmiechem
- (c) powstrzymać się od śmiechu
- (d) nie móc przestać się śmiać
- (e) nie wiem

stifle to try to smother a feeling

17. *To infatuate somebody* oznacza:

- (a) utuczyć kogoś
- (b) ocalić kogoś
- (c) przerazić kogoś
- (d) ogłupić kogoś
- (e) nie wiem

infatuate to make a person utterly fatuous

18. *A scrawny person* to osoba która jest:

- (a) skąpa
- (b) chuda
- (c) otyła
- (d) szczerza
- (e) nie wiem

scrawny (of a person) scraggy

19. *A peevish look* to spojrzenie pełne:

- (a) zdenerwowania
- (b) dobroci
- (c) powagi
- (d) zadumy
- (e) nie wiem

peevish showing vexation

20. *Superfluity* to:

- (a) wspaniałość
- (b) nadmiar czegoś
- (c) płynność
- (d) pozostałość czegoś
- (e) nie wiem

superfluity surfeit (of something)

Task 1, version 2 (COBUILD-8)

Zadanie 1. Przeczytaj definicję wytłuszczonego wyrazu w dolnej części ramki i zakreśl jedną odpowiedź (a, b, c, d), która najbardziej odpowiada znaczeniu tego wyrazu. Jeśli nie wiesz, zaznacz odpowiedź e.

1. **Pungent** smell to:

- (a) przyjemny zapach
- (b) mdły zapach
- (c) drażniący zapach
- (d) ledwo wyczuwalny zapach
- (e) nie wiem

pungent Something that is **pungent** has a strong, sharp smell or taste which is often so strong that it is unpleasant

2. **Loophole** to:

- (a) niechęć pracodawcy do sporządzania umów
- (b) brak przepisów prawnych normujących jakąś sytuację lub błąd w przepisach
- (c) natychmiastowe rozwiązanie umowy z pracownikiem
- (d) środki ostrożności podejmowane przez pracodawców
- (e) nie wiem

loophole ... A **loophole** in the law is a small mistake which allows people to do something that would otherwise be illegal.

3. Distinguished oznacza:

- (a) poważany
- (b) ilustrowany
- (c) apodyktyczny
- (d) mało charakterystyczny
- (e) nie wiem

distinguished ... 1 If you describe a person or their work as **distinguished**, you mean that they have been very successful in their career and have a good reputation.

4. Hillbilly to:

- (a) rustykalny urok
- (b) prostak
- (c) osoba nieodgadniona
- (d) osoba skromna
- (e) nie wiem.

hillbilly ... If you refer to someone as a **hillbilly**, you are saying in a fairly rude way that you think they are uneducated and stupid because they come from the countryside

5. Lopsided to:

- (a) wąski
- (b) znajdujący się przy samym brzegu czegoś
- (c) odwrócony do góry nogami
- (d) pochylony do boku
- (e) nie wiem.

lopsided ... Something that is **lopsided** is uneven because one side is lower or heavier than the other.

6. Hobnob oznacza:

- (a) przebywać z rodziną
- (b) wmieszać się w tłum
- (c) kręcić się koło kogoś, przestawać z kimś
- (d) wstąpić do towarzystwa, klubu lub partii
- (e) nie wiem

hobnob ... If you disapprove of the way in which someone is spending a lot of time with a group of people, especially rich and powerful people, you can say that he or she is **hobnobbing** with them.

7. Stingy with words oznacza:

- (a) używający obraźliwych słów
- (b) wrażliwy na czyjeś słowa
- (c) pazerny na pieniądze
- (d) małomówny
- (e) nie wiem.

stingy ... If you describe someone as stingy, you are criticizing them for being unwilling to spend money.

8. Obfuscate to:

- (a) połączyć
- (b) rozwścieczyć
- (c) zaciekawić
- (d) zaciemnić
- (e) nie wiem

obfuscate ... To **obfuscate** something means to deliberately make it seem confusing and difficult to understand.

9. Loom to:

- (a) szybko następować po sobie, galopować
- (b) zbliżać się wielkimi krokami
- (c) odwołać
- (d) naświetlić
- (e) nie wiem

loom ... If a worrying or threatening situation or event is **looming**, it seems likely to happen soon

10. Abject failure to:

- (a) przypadkowe niepowodzenie
- (b) żalosne, okropne niepowodzenie
- (c) nieuchronne niepowodzenie
- (d) błahe, drobne niepowodzenie
- (e) nie wiem

abject ... You use **abject** to emphasize that a situation or quality is extremely bad

11. Debauchery to:

- (a) bezwzględny egoizm
- (b) namiętne pragnienie
- (c) chciwość
- (d) rozpasanie, rozpusta
- (e) nie wiem

debauchery the habit of indulging in dissolute practices

12. *To redeem a mistake* oznacza:

- (a) uniknąć błędu
- (b) naprawić błąd
- (c) skrytykować błąd
- (d) doszukiwać się błędu
- (e) nie wiem

redeem to atone (a mistake)

13. *An invidious topic* to temat:

- (a) który może być nieprzyjemny
- (b) który dotyczy choroby
- (c) który narzuca komuś sposób myślenia
- (d) który trudno jest określić
- (e) nie wiem

invidious likely to entail ill-will

14. *An impudent person* to osoba:

- (a) która stroni od innych ludzi
- (b) która jest poważna
- (c) która jest wyjątkowo uparta
- (d) która jest bezczelna i zuchwała
- (e) nie wiem

impudent showing unblushing insolence to other people

15. *Solitude* to:

- (a) poczucie winy
- (b) solidarność
- (c) odosobnienie
- (d) poczucie bezpieczeństwa
- (e) nie wiem

solitude seclusion

16. *To stifle a smile* oznacza:

- (a) uśmiechnąć się szeroko
- (b) wybuchnąć śmiechem
- (c) powstrzymać się od śmiechu
- (d) nie móc przestać się śmiać
- (e) nie wiem

stifle to try to smother a feeling

17. *To infatuate somebody* oznacza:

- (a) utuczyć kogoś
- (b) ocalić kogoś
- (c) przerazić kogoś
- (d) ogłupić kogoś
- (e) nie wiem

infatuate to make a person utterly fatuous

18. *A scrawny person* to osoba która jest:

- (a) skąpa
- (b) chuda
- (c) otyła
- (d) szczera
- (e) nie wiem

scrawny (of a person) scraggy

19. *A peevish look* to spojrzenie pełne:

- (a) zdenerwowania
- (b) dobroci
- (c) powagi
- (d) zadumy
- (e) nie wiem

peevish showing vexation

20. *Superfluity* to:

- (a) wspaniałość
- (b) nadmiar czegoś
- (c) płynność
- (d) pozostałość czegoś
- (e) nie wiem

superfluity surfeit (of something)

Task 1, version 3 (ODE-2)

Zadanie I. Przeczytaj definicję wytłuszczonego wyrazu w dolnej części ramki i zakreśl jedną odpowiedź (a, b, c, d), która najbardziej odpowiada znaczeniu tego wyrazu. Jeśli nie wiesz, zaznacz odpowiedź e.

1. Pungent smell to:

- (a) przyjemny zapach
- (b) mdły zapach
- (c) drażniący zapach
- (d) ledwo wyczuwalny zapach
- (e) nie wiem

pungent having a sharply strong taste or smell

2. Loophole to:

- (a) niechęć pracodawcy do sporządzania umów
- (b) brak przepisów prawnych normujących jakąś sytuację lub błąd w przepisach
- (c) natychmiastowe rozwiązanie umowy z pracownikiem
- (d) środki ostrożności podejmowane przez pracodawców
- (e) nie wiem

loophole an ambiguity or inadequacy in the law or a set of rules

3. Distinguished oznacza:

- (a) poważany
- (b) ilustrowany
- (c) apodyktyczny
- (d) mało charakterystyczny
- (e) nie wiem

distinguished very successful, authoritative, and commanding great respect

4. Hillbilly to:

- (a) rustykalny urok
- (b) prostak
- (c) osoba nieodgadniona
- (d) osoba skromna
- (e) nie wiem.

hillbilly an unsophisticated country person, as associated originally with the remote regions of the Appalachians

5. Lopsided to:

- (a) wąski
- (b) znajdujący się przy samym brzegu czegoś
- (c) odwrócony do góry nogami
- (d) pochylony do boku
- (e) nie wiem.

lopsided with one side lower or smaller than the other

6. Hobnob oznacza:

- (a) przebywać z rodziną
- (b) wmieszać się w tłum
- (c) kręcić się koło kogoś, przestawać z kimś
- (d) wstąpić do towarzystwa, klubu lub partii
- (e) nie wiem

hobnob mix socially, especially with those of perceived higher social status

7. Stingy with words oznacza:

- (a) używający obraźliwych słów
- (b) wrażliwy na czyjeś słowa
- (c) pazerny na pieniądze
- (d) małomówny
- (e) nie wiem.

stingy mean; ungenerous

8. Obfuscate to:

- (a) połączyć
- (b) rozwścieczyć
- (c) zaciekać
- (d) zaciemnić
- (e) nie wiem

obfuscate make obscure, unclear, or unintelligible

9. Loom to:

- (a) szybko następować po sobie, galopować
- (b) zbliżać się wielkimi krokami
- (c) odwołać
- (d) naświetlić
- (e) nie wiem

loom (of an event regarded as threatening) seem about to happen

10. Abject failure to:

- (a) przypadkowe niepowodzenie
- (b) żałosne, okropne niepowodzenie
- (c) nieuchronne niepowodzenie
- (d) błahe, drobne niepowodzenie
- (e) nie wiem

abject (of a situation or condition) extremely unpleasant and degrading

11. *Debauchery* to:

- (a) bezwzględny egoizm
- (b) namiętne pragnienie
- (c) chciwość
- (d) rozpasanie, rozpusta
- (e) nie wiem

debauchery the habit of allowing oneself to behave immorally

12. *To redeem a mistake* oznacza:

- (a) uniknąć błędu.
- (b) naprawić błąd.
- (c) skrytykować błąd
- (d) doszukiwać się błędu
- (e) nie wiem

redeem to compensate for (a mistake)

13. *An invidious topic* to temat:

- (a) który może być nieprzyjemny
- (b) który dotyczy choroby
- (c) który narzuca komuś sposób myślenia
- (d) który trudno jest określić
- (e) nie wiem

invidious likely to cause unkind feelings

14. *An impudent person* to osoba:

- (a) która stroni od innych ludzi
- (b) która jest poważna
- (c) która jest wyjątkowo uparta
- (d) która jest bezczelna i zuchwała
- (e) nie wiem

impudent showing no respect to other people

15. *Solitude* to:

- (a) poczucie winy
- (b) solidarność
- (c) odosobnienie
- (d) poczucie bezpieczeństwa
- (e) nie wiem

solitude loneliness

16. *To stifle a smile* oznacza:

- (a) uśmiechnąć się szeroko
- (b) wybuchnąć śmiechem
- (c) powstrzymać się od śmiechu
- (d) nie móc przestać się śmiać
- (e) nie wiem

stifle to try not to express a feeling

17. *To infatuate somebody* oznacza:

- (a) utuczyć kogoś
- (b) ocalić kogoś
- (c) przerazić kogoś
- (d) oglupić kogoś
- (e) nie wiem

infatuate to make a person very silly

18. *A scrawny person* to osoba która jest:

- (a) skąpa
- (b) chuda
- (c) otyła
- (d) szczera
- (e) nie wiem

scrawny (of a person) very thin

19. *A peevish look* to spojrzenie pełne:

- (a) zdenerwowania
- (b) dobroci
- (c) powagi
- (d) zadumy
- (e) nie wiem

peevish showing irritation

20. *Superfluity* to:

- (a) wspaniałość
- (b) nadmiar czegoś
- (c) płynność
- (d) pozostałość czegoś
- (e) nie wiem

superfluity a larger amount (of something)

Task 1, version 4 (*Chambers-12*)

Zadanie 1. Przeczytaj definicję wytłuszczonego wyrazu w dolnej części ramki i zakreśl jedną odpowiedź (a, b, c, d), która najbardziej odpowiada znaczeniu tego wyrazu. Jeśli nie wiesz, zaznacz odpowiedź e.

1. **Pungent** smell to:

- (a) przyjemny zapach
- (b) mdły zapach
- (c) drażniący zapach (d) ledwo wyczuwalny zapach (e) nie wiem

pungent pricking or acrid in taste or smell

2. **Loophole** to:

- (a) niechęć pracodawcy do sporządzania umów
- (b) brak przepisów prawnych normujących jakąś sytuację lub błąd w przepisach
- (c) natychmiastowe rozwiązanie umowy z pracownikiem
- (d) środki ostrożności podejmowane przez pracodawców
- (e) nie wiem

loophole a means of evasion, for example an ambiguity in a contract, etc.

3. **Distinguished** oznacza:

- (a) poważany
- (b) ilustrowany
- (c) apodyktyczny
- (d) mało charakterystyczny
- (e) nie wiem

distinguished illustrious, eminent

4. **Hillbilly** to:

- (a) rustykalny urok
- (b) prostak
- (c) osoba nieodgadniona
- (d) osoba skromna
- (e) nie wiem.

hillbilly a rustic of the hill country; any unsophisticated person

5. Lopsided to:

- (a) wąski
- (b) znajdujący się przy samym brzegu czegoś
- (c) odwrócony do góry nogami
- (d) pochylony do boku
- (e) nie wiem.

lopsided leaning to one side, off balance

6. Hobnob oznacza:

- (a) przebywać z rodziną
- (b) wmieszać się w tłum
- (c) kręcić się koło kogoś, przestawać z kimś
- (d) wstąpić do towarzystwa, klubu lub partii
- (e) nie wiem

hobnob to associate or drink together familiarly

7. Stingy with words oznacza:

- (a) używający obraźliwych słów
- (b) wrażliwy na czyjeś słowa
- (c) pazerny na pieniądze
- (d) mało mówny
- (e) nie wiem.

stingy niggardly

8. Obfuscate to:

- (a) połączyć
- (b) rozwścieczyć
- (c) zaciekać
- (d) zaciemnić
- (e) nie wiem

obfuscate to obscure; to confuse or bewilder

9. Loom to:

- (a) szybko następować po sobie, galopować
- (b) zbliżać się wielkimi krokami
- (c) odwołać
- (d) naświetlić
- (e) nie wiem

loom (of an event) to impend, be imminent

10. Abject failure to:

- (a) przypadkowe niepowodzenie
- (b) żalosne, okropne niepowodzenie
- (c) nieuchronne niepowodzenie
- (d) błahe, drobne niepowodzenie
- (e) nie wiem

abject mean, wretched

11. Debauchery to:

- (a) bezwzględny egoizm
- (b) namiętne pragnienie
- (c) chciwość
- (d) rozpasanie, rozpusta
- (e) nie wiem

debauchery the habit of allowing oneself to behave immorally

12. To redeem a mistake oznacza:

- (a) uniknąć błędu.
- (b) naprawić błąd.
- (c) skrytykować błąd
- (d) doszukiwać się błędu
- (e) nie wiem

redeem to compensate for (a mistake)

13. An invidious topic to temat:

- (a) który może być nieprzyjemny
- (b) który dotyczy choroby
- (c) który narzuca komuś sposób myślenia
- (d) który trudno jest określić
- (e) nie wiem

invidious likely to cause unkind feelings

14. An impudent person to osoba:

- (a) która stroni od innych ludzi
- (b) która jest poważna
- (c) która jest wyjątkowo uparta
- (d) która jest bezczelna i zuchwała
- (e) nie wiem

impudent showing no respect to other people

15. *Solitude* to:

- (a) poczucie winy
- (b) solidarność
- (c) odosobnienie
- (d) poczucie bezpieczeństwa
- (e) nie wiem

solitude loneliness

16. *To stifle a smile* oznacza:

- (a) uśmiechnąć się szeroko
- (b) wybuchnąć śmiechem
- (c) powstrzymać się od śmiechu
- (d) nie móc przestać się śmiać
- (e) nie wiem

stifle to try not to express a feeling

17. *To infatuate somebody* oznacza:

- (a) utuczyć kogoś
- (b) ocalić kogoś
- (c) przerazić kogoś
- (d) oğłupić kogoś
- (e) nie wiem

infatuate to make a person very silly

18. *A scrawny person* to osoba, która jest:

- (a) skąpa
- (b) chuda
- (c) otyła
- (d) szczerą
- (e) nie wiem

scrawny (of a person) very thin

19. *A peevish look* to spojrzenie pełne:

- (a) zdenerwowania
- (b) dobroci
- (c) powagi
- (d) zadumy
- (e) nie wiem

peevish showing irritation

20. <i>Superfluity</i> to: (a) wspaniałość (b) nadmiar czegoś (c) płynność (d) pozostałość czegoś (e) nie wiem
<i>superfluity</i> a larger amount (of something)

Task 2, version 1 (all the dictionaries)

Zadanie II. Zaznacz w skali 1-4, w jakim stopniu definicja jest pomocna w zrozumieniu znaczenia słowa.⁵

1 = w ogóle nie jest pomocna

4 = bardzo pomocna

can (n.)	can (n.)	can (n.)	can (n.)
a metal container in which food or drink is preserved without air	a small steel or aluminium container in which food or drink is hermetically sealed for storage over long periods	A can is a metal container in which something such as food, drink, or paint is put. The container is usually sealed to keep the contents fresh	a tin, a container of tin-plate in which meat, fruit, etc are sealed up to preserve them
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

stingy (adj.)	stingy (adj.)	stingy (adj.)	stingy (adj.)
not generous, especially with money	If you describe someone as stingy, you are criticizing them for being unwilling to spend money.	mean; ungenerous	niggardly
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

5. [Indicate how helpful each definition is in understanding the meaning of the word. Use the scale from 1 to 4. 1 = “unhelpful” 4 = “very helpful”]

quandary (n.)	quandary (n.)	quandary (n.)	quandary (n.)
a dilemma	If you are in a quandary, you have to make a decision but cannot decide what to do.	a difficult situation or problem, especially one in which you cannot decide what to do	a difficult situation
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

embarrass (v.)	embarrass (v.)	embarrass (v.)	embarrass (v.)
to make someone feel ashamed, nervous, or uncomfortable, especially in front of other people	If something or someone embarrasses you, they make you feel shy or ashamed.	cause (someone) to feel awkward, self-conscious, or ashamed	to cause to feel self-conscious, ashamed or awkward
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

dimple (n.)	dimple (n.)	dimple (n.)	dimple (n.)
A dimple is a small hollow in someone's cheek or chin, often one that you can see when they smile.	a small depression in the flesh, either one that exists permanently or one that forms in the cheeks when one smiles	a small hollow, especially on a person's cheek or chin	a small hollow place on your skin, especially one on your cheek or chin when you smile
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

steppe (n.)	steppe (n.)	steppe (n.)	steppe (n.)
dry, grassy, generally treeless, uncultivated and sometimes salt plain, as in central Europe and in Asia	Steppes are large areas of flat grassy land where there are no trees, especially the area that stretches from Eastern Europe across the south of the former Soviet Union to Siberia.	a large areas of land without trees, especially in Russia, Asia, and eastern Europe	a large area of flat unforested grassland in SE Europe or Siberia
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

powder(n.)	powder(n.)	powder(n.)	powder(n.)
any solid in the form of fine dust-like particles	fine, dry particles produced by the grinding, crushing, or disintegration of a solid substance	Powder consists of many tiny particles of a solid substance	a dry substance in the form of very small grains
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

inscrutable	inscrutable	inscrutable	inscrutable
someone who is inscrutable shows no emotion or reaction in the expression on their face so that it is impossible to know what they are feeling or thinking	that cannot be scrutinized or searched into and understood	impossible to understand or interpret	If a person or their expression is inscrutable , it is very hard to know what they are really thinking or what they mean.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

trumpet (n.)	trumpet (n.)	trumpet (n.)	trumpet (n.)
an orchestral, military and signalling brass instrument of powerful and brilliant tone, in its present form a narrow tube bent twice upon itself, with cupped mouthpiece and flaring bell, giving, by action of the lips and breath pressure, harmonics of its fundamental, the scale filled up by use of crooks, slides, or valves	A trumpet is a musical instrument of the brass family which plays quite high notes.	a musical instrument that you blow into, which consists of a curved metal tube that is wide at the end, and three buttons you press to change the notes	a brass musical instrument with a flared bell and a bright, penetrating tone. The modern instrument has the tubing looped to form a straight-sided coil, with three valves.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

abject (adj.)	abject (adj.)	abject (adj.)	abject (adj.)
the state of being extremely poor, unhappy, unsuccessful etc	(of a situation or condition) extremely unpleasant and degrading	You use abject to emphasize that a situation or quality is extremely bad	mean, wretched
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

embarrassment (n.)	embarrassment (n.)	embarrassment (n.)	embarrassment (n.)
the state of feeling embarrassed	a feeling of self-consciousness, shame, or awkwardness	the feeling you have when you are embarrassed	Embarrassment is the feeling you have when you are embarrassed
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

hard (adj)	hard (adj)	hard (adj)	hard (adj)
hard water contains a lot of minerals, and does not mix easily with soap	Hard water contains a lot of calcium compounds that stop soap making bubbles and sometimes appear as a deposit in kettles and baths.	(of water) containing mineral salts that make lathering difficult.	(of water) difficult to lather owing to calcium or magnesium salt in solution
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

potential(adj.)	potential(adj.)	potential(adj.)	potential(adj.)
You use potential to say that someone or something is capable of developing into the particular kind of person or thing mentioned.	possible or likely, though as yet not tested or actual	having or showing the capacity to develop into something in the future	likely to develop into a particular type of person or thing in the future
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

ski (v.)	ski (v.)	ski (v.)	ski (v.)
When people ski , they move over snow or water on skis.	travel over snow on skis	to move on skis for sport or in order to travel on snow or water	to travel on skis
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

loom (v.)	loom (v.)	loom (v.)	loom (v.)
If a worrying or threatening situation or event is looming , it seems likely to happen soon	(of an event) to impend, be imminent	if a problem or difficulty looms, it is likely to happen very soon	(of an event regarded as threatening) seem about to happen
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

lopsided (adj.)	lopsided (adj.)	lopsided (adj.)	lopsided (adj.)
with one side lower or smaller than the other	leaning to one side, off balance	having one side that is lower or heavier than the other	Something that is lopsided is uneven because one side is lower or heavier than the other.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

rattan	rattan	rattan	rattan
the thin jointed stems of a palm, used to make furniture.	any of various climbing palms, especially of the genus Calamus, with a very long thin stem; ... the stems collectively as wickerwork	the plant used to make WICKER furniture	Rattan furniture is made from the woven strips of stems of a plant which grows in South East Asia.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

hobnob (v.)	hobnob (v.)	hobnob (v.)	hobnob (v.)
If you disapprove of the way in which someone is spending a lot of time with a group of people, especially rich and powerful people, you can say that he or she is hobnobbing with them.	to associate or drink together familiarly	to spend time talking to people who are in a higher social position than you	mix socially, especially with those of perceived higher social status
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

insult (v.)	insult (v.)	insult (v.)	insult (v.)
speak to or treat with disrespect or scornful abuse	to treat with indignity or contempt	to offend someone by saying or doing something they think is rude	If someone insults you, they say or do something that is rude or offensive.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

hillbilly (n.)	hillbilly (n.)	hillbilly (n.)	hillbilly (n.)
a rustic of the hill country; any unsophisticated person	an unsophisticated country person, as associated originally with the remote regions of the Appalachians	an insulting word meaning an uneducated poor person who lives in the mountains	If you refer to someone as a hillbilly , you are saying in a fairly rude way that you think they are uneducated and stupid because they come from the countryside
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

feather (n.)	feather (n.)	feather (n.)	feather (n.)
any of the flat appendages growing from a bird's skin and forming its plumage, consisting of a partly hollow horny shaft fringed with vanes of barbs	one of the light soft things that cover a bird's body	one of the light growths that form the soft covering of a bird	A bird's feathers are the soft covering on its body. Each feather consists of a lot of smooth hairs on each side of a thin stiff centre.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

redeem (v.)	redeem (v.)	redeem (v.)	redeem (v.)
atone or make amends for (sin, error, or evil)	to atone for; to compensate for	If you redeem yourself or your reputation, you do something that makes people have a good opinion of you again after you have behaved or performed badly.	to do something that will improve what other people think of you, after you have behaved badly or failed
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

ski (n.)	ski (n.)	ski (n.)	ski (n.)
each of a pair of long narrow pieces of hard, flexible material, typically pointed and turned up at the front, fastened under the feet for travelling over snow	a long narrow runner originally of wood, now also of metal or synthetic materials, fastened to the foot to enable the wearer to slide across snow, etc	Skis are long, flat, narrow pieces of wood, metal, or plastic that are fastened to boots so that you can move easily on snow or water.	one of a pair of long thin narrow pieces of wood or plastic that you fasten to your boots and use for moving on snow or on water
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

accuse (v.)	accuse (v.)	accuse (v.)	accuse (v.)
to bring a charge against	to say that you believe someone is guilty of a crime or of doing something bad	If you are accused of a crime, a witness or someone in authority states or claims that you did it, and you may be formally charged with it and put on trial.	charge (someone) with an offence or crime
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

rancid (adj.)	rancid (adj.)	rancid (adj.)	rancid (adj.)
If butter, bacon, or other oily foods are rancid , they have gone bad and taste old and unpleasant.	oily or fatty food that is rancid smells or tastes unpleasant because it is no longer fresh	rank in smell or taste, as of butter or oil that is going bad	(of foods containing fat or oil) smelling or tasting unpleasant as a result of being old and stale
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

Task 2, version 2 (all the dictionaries)

Zadanie II. Zaznacz w skali 1-4, w jakim stopniu definicja jest pomocna w zrozumieniu znaczenia słowa.

1 = w ogóle nie jest pomocna

4 = bardzo pomocna

gherkin	gherkin	gherkin	gherkin
a small type of CUCUMBER that has been preserved in VINEGAR to make a type of PICKLE	a small cucumber used for pickling	Gherkins are small green cucumbers that have been preserved in vinegar.	a small variety of cucumber, or a young green cucumber used for pickling
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

pungent (adj.)	pungent (adj.)	pungent (adj.)	pungent (adj.)
having a sharply strong taste or smell	pricking or acrid in taste or smell	having a strong taste or smell	Something that is pungent has a strong, sharp smell or taste which is often so strong that it is unpleasant
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

skill (n.)	skill (n.)	skill (n.)	skill (n.)
an ability to do something well, especially because you have learned and practised it	the ability to do something well	expertness	Skill is the knowledge and ability that enables you to do something well.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

loophole (n.)	loophole (n.)	loophole (n.)	loophole (n.)
A loophole in the law is a small mistake which allows people to do something that would otherwise be illegal.	a small mistake in a law that makes it possible to avoid doing something that the law is supposed to make you do	a means of evasion, for example an ambiguity in a contract, etc.	an ambiguity or inadequacy in the law or a set of rules
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

wrench (n.)	wrench (n.)	wrench (n.)	wrench (n.)
a metal tool that you use for turning NUTS ... spanner	an adjustable tool like a spanner, used for gripping and turning nuts or bolts	an instrument for turning nuts, etc, an adjustable spanner	A wrench or a monkey wrench is an adjustable metal tool used for tightening or loosening metal nuts of different sizes.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

hiccup (n.)	hiccup (n.)	hiccup (n.)	hiccup (n.)
an involuntary spasm of the diaphragm and respiratory organs, with a sudden closure of the glottis and a characteristic gulping sound	the involuntary contraction of the diaphragm while the glottis is spasmodically closed	a sudden repeated stopping of the breath, usually caused by eating or drinking too fast	When you have hiccups , you make repeated sharp sounds in your throat, often because you have been eating or drinking too quickly.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

parrot(n.)	parrot(n.)	parrot(n.)	parrot(n.)
one of a family of tropical and subtropical birds with brilliant plumage, a hooked bill, and zygodactyl feet, good imitators of human speech	A parrot is a tropical bird with a curved beak and brightly-coloured or grey feathers. Parrots can be kept as pets. Some parrots are able to copy what people say.	a bird, often vividly coloured, with a short downcurved hooked bill, grasping feet, and a raucous voice, found especially in the tropics and feeding on fruits and seeds. Many are popular as cage birds, and some are able to mimic the human voice.	a tropical bird with a curved beak and brightly coloured feathers that can be taught to copy human speech
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

grout (n.)	grout (n.)	grout (n.)	grout (n.)
a mortar or paste for filling crevices, especially the gaps between wall or floor tiles	a thin coarse mortar for filling cracks, etc	a mixture of sand, water and CEMENT or LIME that you spread between TILES when you fix them to a wall	Grout is a thin mixture of sand, water, and cement or lime, which is used to fill in the spaces between tiles that are fixed to a wall.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

distinguished (adj.)	distinguished (adj.)	distinguished (adj.)	distinguished (adj.)
very successful, authoritative, and commanding great respect	successful, respected, and admired	illustrious, eminent	1 If you describe a person or their work as distinguished , you mean that they have been very successful in their career and have a good reputation.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

instigate (v.)	instigate (v.)	instigate (v.)	instigate (v.)
to make a process start, especially one relating to law or politics	Someone who instigates an event causes it to happen.	bring about or initiate (an action or event)	to initiate, bring about
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

horde (n.)	horde (n.)	horde (n.)	horde (n.)
a large crowd moving in a noisy uncontrolled way	If you describe a crowd of people as a horde , you mean that the crowd is very large and excited and, often, rather frightening or unpleasant	1 ... a large group of people	a multitude
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

compulsory (adj.)	compulsory (adj.)	compulsory (adj.)	compulsory (adj.)
something that is compulsory must be done because it is the law or because someone in authority orders you to	required by law or a rule	If something is compulsory , you must do it or accept it, because it is the law or because someone in a position of authority says you must.	obligatory
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

pulsar (n.)	pulsar (n.)	pulsar (n.)	pulsar (n.)
an object like a star that is far away in space and produces RADIATION and RADIO WAVES	any of a number of interstellar sources of regularly pulsed radiation, first discovered in 1967, and thought to be rotating neutron stars	a celestial object, thought to be a rapidly rotating neutron star, that emits regular pulses of radio waves and other electromagnetic radiation at rates of up to one thousand pulses per second	A pulsar is a star that spins very fast and cannot be seen but produces regular radio signals.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

trudge (v.)	trudge (v.)	trudge (v.)	trudge (v.)
to walk with labour or effort	walk slowly and with heavy steps, typically because of exhaustion or harsh conditions	to walk with slow heavy steps, especially because you are tired or it is difficult to walk	If you trudge somewhere, you walk there slowly and with heavy steps, especially because you are tired or unhappy.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

solitude (n.)	solitude (n.)	solitude (n.)	solitude (n.)
when you are alone, especially when this is what you enjoy	Solitude is the state of being alone, especially when this is peaceful and pleasant.	the state or situation of being alone	solitariness
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

stipend (n.)	stipend (n.)	stipend (n.)	stipend (n.)
an amount of money paid regularly to someone, especially a priest, as a salary or as money to live on	a salary, especially a Scottish parish minister's	a fixed regular sum paid as a salary or as expenses to a clergyman, teacher, or public official	A stipend is a sum of money that is paid regularly, especially to a magistrate or a member of the clergy, as a salary or for their living expenses
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

persuade	persuade	persuade	persuade
induce (someone) to do something through reasoning or argument	to induce by reasoning, advice, etc	If you persuade someone to do something, you cause them to do it by giving them good reasons for doing it.	to make someone decide to do something, especially by giving them reasons why they should do it, or asking them many times to do it
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

permission	permission	permission	permission
an act of permitting	If someone who has authority over you gives you permission to do something, they say that they will allow you to do it.	if you have permission to do something, you are officially allowed to do it	the action of officially allowing someone to do a particular thing
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

obfuscate	obfuscate	obfuscate	obfuscate
to deliberately make something unclear or difficult to understand	to obscure; to confuse or bewilder	make obscure, unclear, or unintelligible	To obfuscate something means to deliberately make it seem confusing and difficult to understand.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

hormone (n.)	hormone (n.)	hormone (n.)	hormone (n.)
an internal secretion that, on reaching some part of a plant or animal body, exercises a specific physiological action	A hormone is a chemical, usually occurring naturally in your body, that makes an organ of your body do something.	a chemical substance produced by your body that influences its growth, development, and condition	a regulatory substance produced in an organism and transported in tissue fluids such as blood or sap to stimulate specific cells or tissues into action
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

embarrassed (adj.)	embarrassed (adj.)	embarrassed (adj.)	embarrassed (adj.)
A person who is embarrassed feels shy, ashamed, or guilty about something.	self-conscious, ashamed or awkward	feeling uncomfortable or nervous and worrying about what people think of you, for example because you have made a silly mistake, or because you have to talk or sing in public	feeling or showing embarrassment
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

stifle (v.)	stifle (v.)	stifle (v.)	stifle (v.)
restrain (a reaction) or stop oneself acting on (an emotion)	to suppress	to stop a feeling from being expressed	If you stifle your natural feelings or behaviour, you prevent yourself from having those feelings or behaving in that way.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

hire (v.)	hire (v.)	hire (v.)	hire (v.)
to procure the use or service of, at a price	to pay money to borrow something for a short period of time	If you hire something, you pay money to the owner so that you can use it for a period of time.	obtain the temporary use of (something) for an agreed payment
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

inquisitive	inquisitive	inquisitive	inquisitive
unduly curious about the affairs of others	asking too many questions and trying to find out too many details about something or someone	apt to ask questions, especially about other people's affairs	An inquisitive person likes finding out about things, especially secret things.
1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

References

Dictionaries

Printed dictionaries

- Allen, Robert E. 1990. *The Concise Oxford Dictionary of Current English*. Oxford: Oxford University Press. (COD8)
- Anderson, Sandra, Lynda Carey, Kay Cullen, Serenella Flackett, Alice Grandison, eds. 1998. *The Chambers Dictionary*. Edinburgh: Chambers Harrap Publishers. (Chambers9)
- Bailey, Nathaniel. 1721. *An Universal Etymological English Dictionary*. London.
- Bańko, Mirosław, ed. 2000. *Inny Słownik Języka Polskiego PWN*. [A Different Dictionary of the Polish Language PWN]. Vol. 1–2. Warsaw: Wydawnictwo Naukowe PWN.
- Berube, Margery S. 1982. *The American Heritage Dictionary*. Second College Edition. Boston: Houghton Mifflin Co. (AHD)
- Brookes, Ian, ed. 2003. *The Chambers Dictionary*. New 9th edition. – Edinburgh: Chambers Harrap Publishers. (Chambers 9new)
- Bullokar, John. 1616. *An English Expositor*. London.
- Cawdrey, Robert. 1604. *A Table Alphabeticall*. London.
- Cockeram, Henry. 1623. *The English Dictionarie: Or, an Interpreter of Hard English Words*. London.
- Cowie, Anthony P., ed. 1989. *Oxford Advanced Learner's Dictionary of Current English*, 4th edition. Oxford: Oxford University Press. (OALD4)
- Crowther, Jonathan, ed. 1995. *Oxford Advanced Learner's Dictionary of Current English*, 5th edition. Oxford: Oxford University Press. (OALD5)
- Davidson, Thomas, ed. 1901. *Chambers's Twentieth Century Dictionary of the English Language*. Edinburgh: W & R. Chambers.
- Deuter, Margaret, Jennifer Bradbery, Joanna Turnbull, eds. 2015. *Oxford Advanced Learner's Dictionary of Current English*, 9th edition. Oxford: Oxford University Press. (OALD9)
- Dubisz, Stanisław, ed. 2003. *Uniwersalny Słownik Języka Polskiego*. [Universal Dictionary of the Polish Language] (4 volumes). Warsaw: Wydawnictwo Naukowe PWN.
- Donald, James, ed. 1867. *Chambers's Etymological Dictionary of the English Language*. Edinburgh: W. & R. Chambers. (CED)
- Flood, Walter E., and Michael West. 1952. *An Explaining and Pronouncing Dictionary of Scientific and Technical Words*. Longmans.
- Flood, Walter E., and Michael West. 1953. "Supplementary Scientific and Technical Vocabulary" In *A General Service List of English Words with Semantic Frequencies and a Supplementary Word-List for the Writing of Popular Science and Technology*, edited by Michael West, 583–588. London: Longmans.
- Fowler, Henry W., and Francis G. Fowler. 1911. *The Concise Oxford Dictionary of Current English*. Oxford: Clarendon Press. (COD1)

- Fowler, Henry W., Herbert G. Le Mesurier, and Edward McIntosh. 1934. *The Concise Oxford Dictionary of Current English*. Oxford: University Press. (COD3)
- Funk, Isaac K., ed. 1893–1895. *A Standard Dictionary of the English Language*. New York: Funk & Wagnalls.
- Gay, Heather, Brian O’Kill, Katherine Seed, Janet Whitcut. 1984. *Longman Dictionary of the English Language*. Second Edition. Harlow: Longman. (LDEL)
- Gove, Phillip B., ed. 1961. *Webster’s Third New International Dictionary of the English Language Unabridged*. Springfield: G. & C. Merriam. (Third New International)
- Guralnik, David. 1972. *Webster’s New World Dictionary of the American Language: Compact School and Office Edition*. New York: World Publishing. (WNWD)
- Hanks, Patrick. 1979. *Collins English Dictionary*. London: Collins. (CoED)
- Hornby, Albert Sydney, and Anthony P. Cowie, eds. 1974. *Oxford Advanced Learner’s Dictionary of Current English*, 3rd edition. Oxford: Oxford University Press. (OALD3)
- Hornby, Albert Sydney, Edward V. Gatenby, Harold Wakefield, eds. 1942. *Idiomatic and Syntactic English Dictionary*. Tokyo: Kaitakusha. (ISED)
- Hornby, Albert Sydney, Edward V. Gatenby, Harold Wakefield, eds. 1948. *A Learner’s Dictionary of Current English*, 1st edition. Oxford: Oxford University Press. (ALD1)
- Hornby, Albert Sydney, Edward V. Gatenby, Harold Wakefield, eds. 1963. *The Advanced Learner’s Dictionary of Current English*, 2nd edition. London: Oxford University Press. (ALD2)
- Hornby, Albert S., and Eric C. Parnwell. 1952a. *The Progressive English Dictionary*. London: Oxford University Press.
- Hornby, Albert S., and Eric C. Parnwell. 1952b. *An English Reader’s Dictionary*. London: Oxford University Press.
- Ilson, Robert. 1983. “Etymological Information: Can it Help our Students?” *ELT Journal* 37, no. 1: 76–82. <https://doi.org/10.1093/elt/37.1.76>
- Ilson, Robert. 1984. *Reader’s Digest Great Illustrated Dictionary*. 2 vols. London: Reader’s Digest. (GID)
- Johnson, Samuel. 1755. *A Dictionary of the English Language*. London: W. Strahan. (Johnson)
- Kaeding, Friedrich Wilhelm. 1898. *Häufigkeitwörterbuch der Deutschen Sprache*. E.S. Mittler & Sohn: Selbstverlag des Herausgebers.
- Kersey, John. 1708. *Dictionarium Anglo-Britannicum: Or a General English Dictionary*. London.
- Kirkpatrick, Elizabeth. 1980. *The Chambers Universal Learners’ Dictionary: International Students’ Edition*. Edinburgh: W & R Chambers. (CULD)
- Marr, Vivian, ed. 2011. *The Chambers Dictionary*. 12th edition. Edinburgh: Chambers Harrap Publishers. (Chambers12)
- Marvick, Karen C. et al. 2014. *Longman Dictionary of Contemporary English*, 6th edition. Harlow: Pearson Longman. (LDOCE6)
- Mayor, Michael, ed. 2009. *Longman Dictionary of Contemporary English*, 5th edition. Harlow: Pearson Longman. (LDOCE5)
- McIntosh, Colin, ed. 2013. *Cambridge Advanced Learner’s Dictionary*. 4th edition. Cambridge: Cambridge University Press. (CALD4)
- Mish, Frederick. 1983. *Webster’s Ninth New Collegiate Dictionary*. 9th edition. Springfield: Merriam-Webster Inc. (Collegiate9)
- Mish, Frederick. 1993. *Merriam-Webster’s Collegiate Dictionary*. 10th edition. Springfield: Merriam-Webster Inc. (Collegiate10)
- Mish, Frederick. 2003. *Merriam-Webster’s Collegiate Dictionary*. 11th edition. Springfield: Merriam-Webster Inc. (Collegiate11)

- Mosiółek-Kłosińska, Katarzyna, and Anna Ciesielska. 2001. *W kilku słowach. Słownik frazeologiczny języka polskiego*. [In a Few Words. A Phraseological Dictionary of Polish Idioms] Warsaw: Wydawnictwo Szkolne PWN.
- Murray, James A. H. 1888. "Preface to Volume I. A and B" In *A New English Dictionary on Historical Principles; Founded Mainly on the Materials Collected by the Philological Society*, edited by James A. H. Murray et al. Oxford: The Clarendon Press.
- Murray, James A. H., Henry Bradley, William A. Craigie, Charles T. Onions. 1933. *The Oxford English Dictionary, Being a Corrected Re-issue with an Introduction, Supplement, and Bibliography of A New English Dictionary on Historical Principles*. Oxford: The Clarendon Press. (OED)
- Pawelec, Radosław. 1999. *Wieża Babel. Słownik Wyrazów Obcych Nie Tylko Dla Gimnazjalisty*. [Tower of Babel. A Dictionary of Borrowed Words, Not Only for Secondary School Pupils]. Warsaw: Wydawnictwo Szkolne PWN.
- Pearsall, Judy, ed. 1998. *The New Oxford Dictionary of English*. Oxford: Oxford University Press. (ODE1/NODE)
- Pearsall, Judy. 1999. *Concise Oxford English Dictionary*. Oxford: University Press. (COD10)
- Pearsall, Judy, and Patrick Hanks, eds. 2003. *Oxford Dictionary of English*. 2nd edition. Oxford: Oxford University Press. (ODE2)
- Pearsall, Judy, and Bill Trumble. 1996. *Oxford English Reference Dictionary*. 2nd edition. Oxford: Oxford University Press. (OERD)
- Perrault, Stephen J. 2008. *Merriam-Webster's Advanced Learner's English Dictionary*, 1st edition. Springfield: Merriam-Webster. (MWALED)
- Phillips, Edward. 1671 [1658]. *The New World of English Words*, 3rd edition. London.
- Procter, Paul, ed. 1978. *Longman Dictionary of Contemporary English*, 1st edition. Harlow: Longman. (LDOCE1)
- Procter, Paul, ed. 1995. *Cambridge International Dictionary of English*. 1st edition. Cambridge: Cambridge University Press. (CALD1/CIDE)
- Richards, Jack C. and Richard Schmidt. 2010. *Longman Dictionary of Language Teaching and Applied Linguistics*. 4th ed. Harlow: Pearson Education Limited.
- Rundell, Michael, ed. 2002. *Macmillan English Dictionary for Advanced Learners*. 1st edition. Oxford: Macmillan Education. (MEDAL1)
- Rundell, Michael, ed. 2007. *Macmillan English Dictionary for Advanced Learners*. 2nd edition. Oxford: Macmillan Education. (MEDAL2)
- Schwarz, Catherine, et al. eds. 1988. *Chambers English Dictionary*. 7th edition. Edinburgh: W & R Chambers. (Chambers7)
- Schwarz, Catherine, et al. eds. 1993. *The Chambers Dictionary*. Edinburgh: Chambers Harrap Publishers. (Chambers8)
- Schwarz, Catherine, ed. 1998. *The Chambers Dictionary*. Edinburgh: Chambers Harrap Publishers. (Chambers9)
- Sinclair, John, ed. 1987. *Collins COBUILD English Language Dictionary*. 1st edition. London and Glasgow: HarperCollins. (COBUILD1)
- Sinclair, John, ed. 1988. *Collins Cobuild Essential English Dictionary*. London: Collins. (CCEED)
- Sinclair, John, ed. 1995. *Collins COBUILD English Dictionary*. 2nd edition. London: HarperCollins. (COBUILD2)
- Sinclair, John, ed. 2001. *Collins COBUILD English Dictionary for Advanced Learners*. 3rd edition. London: HarperCollins. (COBUILD3)
- Sinclair, John, ed. 2003. *Collins COBUILD Advanced Learner's English Dictionary*. 4th edition. Glasgow: HarperCollins. (COBUILD4)

- Sinclair, John, ed. 2006. *Collins COBUILD Advanced Learner's English Dictionary*. 5th edition. London and Glasgow: HarperCollins. (COBUILD5)
- Sinclair, John, ed. 2009. *Collins COBUILD Advanced Dictionary*. 6th edition. London and Glasgow: HarperCollins. (COBUILD6)
- Sinclair, John, ed. 2012. *Collins COBUILD Advanced Dictionary of English*. 7th edition. London and Glasgow: HarperCollins. (COBUILD7)
- Sinclair, J. ed. 2014. *Collins COBUILD Advanced Learner's Dictionary*. 8th edition. London and Glasgow: HarperCollins. (COBUILD8)
- Stevenson, Angus, and Maurice Waite. 2011. *Concise Oxford English Dictionary*. 12th edition. Oxford: Oxford University Press. (COD12)
- Summers, Della, ed. 1995. *Longman Dictionary of Contemporary English*, 3rd edition. Harlow: Longman. (LDOCE3)
- Summers, Della, ed. 2003. *Longman Dictionary of Contemporary English*, 4th edition. Harlow: Longman. (LDOCE4)
- Summers, Della, and Michael Rundell, eds. 1987. *Longman Dictionary of Contemporary English*, 2nd edition. Harlow: Longman. (LDOCE2)
- Sykes, John B. 1976. *The Concise Oxford Dictionary of Current English*. Oxford: University Press. (COD6)
- Sykes, John B. 1982. *The Concise Oxford Dictionary of Current English*. Oxford: University Press. (COD7)
- Thompson, Della. 1995. *The Concise Oxford Dictionary of Current English*. Oxford: University Press. (COD9)
- Turnbull, Joanna, ed. 2010. *Oxford Advanced Learner's Dictionary of Current English*, 8th edition. Oxford: Oxford University Press. (OALD8)
- Urdang, Laurence. 1966. *The Random House Dictionary of the English Language*. New York: Random House.
- Walter, Elizabeth, ed. 2005. *Cambridge Advanced Learner's Dictionary*. 2nd edition. Cambridge: Cambridge University Press. (CALD2)
- Walter, Elizabeth, ed. 2008. *Cambridge Advanced Learner's Dictionary*. 3rd edition. Cambridge: Cambridge University Press. (CALD3)
- Webster, Noah. 1864. *An American Dictionary of the English Language*. Royal quarto edition. Revised by Chauncey A. Goodrich and Noah Porter. Springfield: G.&C. Merriam. (Webster-Mahn)
- Wehmeier, Sally, ed. 2000. *Oxford Advanced Learner's Dictionary of Current English*, 6th edition. Oxford: Oxford University Press. (OALD6)
- Wehmeier, Sally, ed. 2005. *Oxford Advanced Learner's Dictionary of Current English*, 7th edition. Oxford: Oxford University Press. (OALD7)
- West, Michael, and James G. Endicott, eds. 1956 [1935]. *The New Method English Dictionary*. London: Longmans, Green. (NMED)
- West, Michael. 1965. *An International Reader's Dictionary*. Longman.
- Williams, Alexander M. 1903. *Collins' Graphic English Dictionary*. London.
- Worcester, Joseph Emerson. 1860. *A Dictionary of the English Language*. Boston: Hickling, Swan, and Brewer. (Worcester)

Dictionaries on CD-ROM/DVDs

- Deuter, Margaret, Jennifer Bradbery, Joanna Turnbull, eds. 2014. *Oxford Advanced Learner's Dictionary of Current English*, DVD-ROM, New 9th edition. Oxford: Oxford University Press. (OALD9)

- Mayor, Michael, ed. 2009. *Longman Dictionary of Contemporary English*, DVD-ROM, 5th edition. Harlow: Pearson Longman. (LDOCE5)
- McIntosh, Colin, ed. 2013. *Cambridge Advanced Learner's Dictionary*. DVD-ROM, 4th edition. Cambridge: Cambridge University Press. (CALD4)
- Rundell, Michael, ed. 2007. *Macmillan English Dictionary for Advanced Learners*. CD-ROM. 2nd edition. Oxford: Macmillan Education. (MEDAL2)
- Simpson, John, and Edmund S. C. Weiner, eds. 2009. *Oxford English Dictionary*. Third Edition on CD-ROM. Oxford University Press.
- Sinclair, John, ed. 2003. *Collins COBUILD Advanced Learner's English Dictionary*, CD-ROM, 4th edition. Glasgow: HarperCollins. (COBUILD4)
- Summers, Della, ed. 2005. *Longman Dictionary of Contemporary English*, Writing Assistant Edition CD-ROM. Harlow: Pearson Education Limited. (LDOCE4)
- Walter, Elizabeth, ed. 2008. *Cambridge Advanced Learner's Dictionary*. DVD-ROM, 3rd edition. Cambridge: Cambridge University Press. (CALD3)
- Wehmeier, Sally, ed. 2005. *Oxford Advanced Learner's Dictionary of Current English*, CD-ROM, 7th edition. Oxford: Oxford University Press. (OALD7)

Online dictionaries

- Cambridge Advanced Learner's Dictionary*. Accessed December 10, 2019. <https://dictionary.cambridge.org/pl/dictionary/>
- Collins COBUILD Advanced Dictionary of English*. Accessed December 10, 2019. <https://www.collinsdictionary.com/dictionary/>
- Longman Dictionary of Contemporary English*. Accessed December 10, 2019. <https://www.ldo-ceonline.com/>
- Macmillan English Dictionary for Advanced Learners*. Accessed December 15, 2019. <https://www.macmillandictionary.com/>
- Merriam-Webster's Advanced Learner's English Dictionary*. Accessed December 4, 2019. <https://learnersdictionary.com/>
- Oxford Advanced Learner's Dictionary*. Accessed December 25, 2019. <https://www.oxfordlearnersdictionaries.com/>

Other literature

- Aarts, Bas, April MacMahon. 2006. *The Handbook of English Linguistics*. Oxford: Blackwell. <https://doi.org/10.1002/9780470753002>
- Abecassis, Michaël. 2007. "Is Lexicography Making Progress? On Dictionary Use and Language Learners' Needs." *Lexikos* 17: 247–258.
- Adamska-Sałaciak, Arleta. 2007. Review of Historical Dictionaries and Historical Dictionary Research: Papers from the International Conference on Historical Lexicography and Lexicology, at the University of Leicester, 2002. Tübingen: Max Niemeyer, edited by Julie Coleman and Anne McDermott. *International Journal of Lexicography*, 20 (2): 182–188. <https://doi.org/10.1093/ijl/ecm002>
- Adamska-Sałaciak, Arleta. 2010a. "Examining Equivalence." *International Journal of Lexicography* 23(4): 387–409. <https://doi.org/10.1093/ijl/ecq024>
- Adamska-Sałaciak, Arleta. 2010b. "Why we need bilingual learners' dictionaries." *English Learners' Dictionaries at the DSNA 2009*, edited by Ilan J. Kernerman and Paul Bogaards, 121–37. TelAviv: K Dictionaries.

- Adamska-Sałaciak, Arleta. 2012. "Dictionary Definitions: Problems and Solutions." *Studia Linguistica Universitatis Iagellonicae Cracoviensis 129 Supplementum*: 323–339.
- Adamska-Sałaciak, Arleta. 2014. "Bilingual Lexicography: Translation Dictionaries." *International Handbook of Modern Lexis and Lexicography*, edited by Patrick Hanks and Gilles-Maurice de Schryver, 1–11. Berlin: Springer-Verlag.
https://doi.org/10.1007/978-3-642-45369-4_6-1
- Aisenstadt, Ester. 1979. "Collocability Restrictions in Dictionaries." *Dictionaries and Their Users: Papers from the 1978 B.A.A.L. Seminar on Lexicography*, edited by Reinhard R. K. Hartmann, 71–74. Exeter: University of Exeter.
- Aitchison, Jean. 1986. "The Gobbledygook Syndrome." *English Today: The International Review of the English Language* 5, January–March: 7–9. <https://doi.org/10.1017/S0266078400001619>
- Allan, Keith. 1986. *Linguistic Meaning* Vol. 1. London: Routledge & Kegan Paul.
- Allen, Jeffrey. 1995. *Review of the Caterpillar KANT English-French MT System, Internal Technical Report*, Peoria, IL: Technical Information Department, Caterpillar Inc.
- Allen, Robert. 1996. The Big Four. *English Today*, 12 (2): 41–55.
<https://doi.org/10.1017/S0266078400009020>
- Anthony, Laurence. 2012. *AntWordProfiler (Version 1.4.0) [Computer Software]*. Tokyo, Japan: Waseda University. Accessed October 1, 2019, www.antlab.sci.waseda.ac.jp/
- Apresjan, Jurij. D. 1972. "Definiowanie znaczeń leksykalnych jako zagadnienie semantyki teoretycznej." In *Semantyka i słownik*, edited by Anna Wierzbicka, 39–58. Wrocław: Zakład Narodowy im. Ossolińskich.
- Apresjan, Jurij. D. 1995. *Semantyka leksykalna. Synonimiczne środki języka*. Wrocław: Zakład Narodowy im. Ossolińskich Wydawnictwo.
- (ASD-STE100) 2017. *ASD Simplified Technical English Specification ASD-STE100*. Access October 2, 2018, <http://www.asd-ste100.org/index.html>
- Atkins, Beryl T. S. 1985. "Monolingual and bilingual learners' dictionaries: A comparison." In *Dictionaries, Lexicography and Language Learning*, edited by Robert Ilson, 15–24. Oxford: Pergamon Press.
- Atkins, Beryl T. S., Charles J. Fillmore, and Christopher R. Johnson. 2003. "Lexicographic Relevance: Selecting Information from Corpus Evidence." *International Journal of Lexicography* 16 (3): 251–80. <https://doi.org/10.1093/ijl/16.3.251>
- Atkins, Beryl T. S., Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Ayto, John. 1984. "The Vocabulary of Definitions." In *Theoretische und Praktische Probleme der Lexicographie*, edited by Dieter Götz and Thomas Herbst, Munich: Max Hueber Verlag.
- Baker, Mona. 1992. *In Other Words: A Course Book on Translation*. London: Routledge.
- Bańko, Mirosław. 2001. *Z pogranicza leksykografii i językoznawstwa. Studia o słowniku jednojęzycznym*. Warsaw: Wydział Polonistyki Uniwersytetu Warszawskiego.
- Bańko Mirosław. 2006. "Circulus vitiosus - circulus odiosus. Uwagi o cyrkularności w definicjach słownikowych." *Manuscula linguistica. In honorem Alexandrae Cieslikowa oblata*, edited by Kazimierz Rymut et al., 21–34. Cracow: Instytut Języka Polskiego PAN.
- Bańko, Mirosław. 2010. "The Polish COBUILD and its Influence on Polish Lexicography." In *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*, edited by Gilles-Maurice de Schryver, 323–332. Kampala 2010 (Menha Linguistic Series).

- Bańko, Mirosław, Magdalena Majdak, Maciej Czeszewski. eds.. 2011. *Słowniki dawne i współczesne. Internetowy przewodnik edukacyjny*. Accessed August 20, 2017, <http://www.leksykografia.uw.edu.pl/slowniki/25/sownik-jezyka-polskiego-warszawa-1978-1981>
- Barnhart, Clarence L. 1962. "Problems in Editing Commercial Monolingual Dictionaries." In *Problems in Lexicography*, edited by Fred W. Householder and Sol Saporta, 161–81. Bloomington, Ind.: Indiana University.
- Baroni, Marco. 2006. "Distributions in Text." In *Corpus Linguistics: An International Handbook*, edited by Anke Lüdeling and Merja Kytö, 803–821. Berlin: Mouton de Gruyter.
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi, Eros Zanchetta. 2009. "The WaCky wide web: a collection of very large linguistically processed web-crawled corpora." *Language Resources and Evaluation* 43. Springer Science+Business Media B.V.: 209–226.
- Battenburg, John D. 1994. "Pioneer in English Lexicography for Language Learners: Michael Philip West." *Dictionaries. Journal of The Dictionary Society of North America*. No. 15, Cleveland: Cleveland State University: 132–148.
- Bauer, Laurie and Paul Nation. 1993. "Word Families." *International Journal of Lexicography* 6 (4): 253–279. <https://doi.org/10.1093/ijl/6.4.253>
- Bauman, John. 2018a. *About the General Service List*. Accessed February 24, 2018, <http://jbauman.com/aboutgsl.html#1953>
- Bauman, John. 2018b. *About the University Word List*. Accessed February, 2018, <http://jbauman.com/aboutUWL.html>
- Beekman, John and Callow, John. 1974. *Translating the Word of God*. Grand Rapids: Zondervan.
- Béjoint, Henri. 1981. "The Foreign Student's Use of Monolingual English Dictionaries: A Study of Language Needs and Reference Skills." *Applied Linguistics* 2 (3): 207–222. <https://doi.org/10.1093/applin/2.3.207>
- Béjoint, Henri. 1987. "The Value of the Dictionary in Vocabulary Acquisition." In *The Dictionary and the Language Learner*, edited by Anthony P. Cowie, 97–104. Tübingen: Niemeyer.
- Béjoint, Henri. 2000. *Modern Lexicography. An Introduction*. Oxford: University Press.
- Bell, Huw. 2013. "Core Vocabulary." *The Encyclopedia of Applied Sciences*, edited by Carol A. Chapelle. Blackwell Publishing.
- Bennett, Judith. 1993. "The Future of Plain Legal Language: Australia: A Case Study." Second International Plain English Conference: Plain English Campaign (UK), London, 24–26 May 1993, Sydney: Centre for Plain Legal Language.
- Benjamin, Martin. 2019. *Teach You Backwards*. Accessed October 17, 2020 <https://www.teachyoubackwards.com/user/malangali/>
- de Bessé, B. 1990. "La Définition Terminologique." In *La Définition*, edited by J. Charraud and F. Mazière, 252–262. Paris: Larousse.
- Bhatia, Tej K. 2006. "Bilingualism and Second Language Learning." In *Encyclopedia of Language and Linguistics*, 2nd ed., edited by Keith Brown, Anne Anderson, Laurie Bauer, Margie Berns, Jim Miller, and Graeme Hirst, 16–22. Oxford: Elsevier. <https://doi.org/10.1016/B0-08-044854-2/00620-9>
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan. 1999. *The Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, Douglas, Susan Conrad, Viviana Cortes. 2003. "Lexical bundles in speech and writing: An initial taxonomy." In *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech*, edited by Andrew Wilson, Paul Rayson, Tony McNery, 71–92. Frankfurt/Main: Peter Lang.

- Biber, D., Conrad, S. & Cortes, V. 2004. "If you look at ... : Lexical Bundles in University Teaching and Textbooks." *Applied Linguistics*, 25 (3), 371–405. <https://doi.org/10.1093/applin/25.3.371>
- Biber, D. 2006. *University Language. A Corpus-Based Study of Spoken and Written Registers*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/scl.23>
- Biden, James. 1843. "On Grammalogues. A letter to the editor of the Phonotypic Journal." *The Phonotypic Journal* Nov. 1843, 2.
- Black, Alison. 1986. *The Effects on Comprehension and Memory of Providign Different Types of Definign Information for New Vocabulary: A Report on Two Experiments Conducted for Longman ELT Dictionaries and Reference Division*. Cambridge: MRC Applied Psychology Unit (unpublished internal report).
- Blake, Norman F. 1996. *A History of the English Language*. London: Macmillan Press Ltd. <https://doi.org/10.1007/978-1-349-24954-1>
- Bogaards, Paul. 1991. "Dictionaries pédagogiques et apprentissage du vocabulaire." *Cahiers de Lexicologie*, 59: 93–107.
- Bogaards, Paul. 1996. "Dictionaries for Learners of English." *International Journal of Lexicography* 9 (4): 277–320. <https://doi.org/10.1093/ijl/9.4.277>
- Bogaards, Paul. 1998. "What Type of Words do Language Learners Look Up?" *Using Dictionaries*, edited by Beryl T. Sue Atkins, 151–157. Tübingen: Max Niemeyer Verlag.
- Bogaards, Paul. 2003. "MEDAL: A Fifth Dictionary for Learners of English." *International Journal of Lexicography* 16 (1): 43–55. <https://doi.org/10.1093/ijl/16.1.43>
- Bogaards, Paul. 2008. "Frequency in Learners' Dictionaries." In *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*, edited by Elisenda Bernal and Janet DeCesaris, 1231–1236. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Bogaards, Paul and Willem A. van der Kloot. 2001. "The Use of Grammatical Information in Learners' Dictionaries." *International Journal of Lexicography* 14 (2): 97–121. <https://doi.org/10.1093/ijl/14.2.97>
- Bogaards, Paul, and Willem A. van der Kloot. 2002. "Verb Constructions in Learners' Dictionaries" In *Proceedings of the Tenth EURALEX International Congress (Vol. II)*, edited by Anna Braasch and Claus Povlsen, 747–757. Copenhagen: Center for Sprogteknologi.
- Bogusławski, Andrzej. 1966. *Semantyczne pojęcie liczebnika i jego morfologia w języku rosyjskim*. Wrocław: Ossolineum.
- Bogusławski, Andrzej. 1970. "On Semantic Primitives and Meaningfulness." In *Signs, Language and Culture*, edited by R. J. A. Greimas, R. Jakobson, M. R. Mayenowa, and S. Żółkiewski, 143–52. The Hague: Mouton.
- Bongers, Herman. 1947. *The History and Principles of Vocabulary Control*. Parts 1 and 2. Woerden: Wocopi.
- Brezina, Vaclav, and Dana Gablasova. 2015. "Is There a Core General Vocabulary? Introducing the New General Service List." *Applied Linguistics* 36 (1): 1–22. <https://doi.org/10.1093/applin/amt018>
- Browne, Charles. 2013. "The New General Service List: Celebration 60 Years of Vocabulary Learning." *The Language Teacher*, 37 (4): 13–16.
- Browne, Charles. 2014. "The New General Service List Version 1.01: Getting Better All the Time." *Korea TESOL Journal* 11 (1): 35–50.
- Bullock, David. 2010. "NSM + LDOCE: A Non-Circular Dictionary of English." *International Journal of Lexicography* 24 (2): 226–240. <https://doi.org/10.1093/ijl/ecq035>

- Bullock, David. 2019. *Learn These Words First*. Accessed March 8, 2019. <http://learnthesewords-first.com/about/what-is-a-multi-layer-dictionary.html>
- Bullon, Stephen, and Geoffrey Leech. 2007. "Longman Communication 3000 and the Longman Defining Vocabulary." In *Longman Communication 3000*, edited by Stephen Bullon and Geoffrey Leech, 1–7. Harlow: Pearson Education.
- Burkhanov, Igor. 1998. *Lexicography A Dictionary of Basic Terminology*. Rzeszów: Wydawnictwo Wyższej Szkoły Pedagogicznej.
- Campion, Mary E., and Warwick B. Elley. 1971. *An Academic Vocabulary List*. Wellington: New Zealand Council for Educational Research.
- Carroll, John, Peter Davies, Barry Richman. 1971. *The American Heritage Word Frequency Book*. New York: American Heritage.
- Carter, Ronald. 1986. "Core Vocabulary and Discourse in the Curriculum: A Question of the Subject." *RELC Journal* 17 (1): 52–70. <https://doi.org/10.1177/003368828601700104>
- Carter, Ronald. 2012. *Vocabulary. Applied Linguistic Perspectives*. Second edition. London: Routledge. <https://doi.org/10.4324/9780203124659>
- Carver, Ronald P. 1994. "Percentage of Unknown Vocabulary Words in Text as a Function of the Relative Difficulty of the Text: Implications for instruction." *Journal of Reading Behavior* 26: 413–437. <https://doi.org/10.1080/10862969409547861>
- Cheng, Yong. 2019. *Joint Training for Neural Machine Translation*. Singapore: Springer. <https://doi.org/10.1007/978-981-32-9748-7>
- Chi, Man L. A. 2003. "Teaching Dictionary Skills in the Classroom." In *Lexicography. Critical Concepts*. Vol. I, edited by Reinhard R. K. Hartmann, 355–369. London: Routledge.
- Chlebda, Wojciech. 2010. "W jakim zakresie słownik dwujęzyczny może być źródłem informacji etnolingwistycznej?" In *Etnolingwistyka a leksykografia*, edited by Wojciech Chlebda, 201–208. Opole: Wydawnictwo Uniwersytetu Opolskiego.
- Chruszczewski, Piotr P. 2006. "Ogden, Charles Kay (1889–1957)." In *Encyclopedia of Language and Linguistics*, 2nd ed., edited by Keith Brown, Anne Anderson, Laurie Bauer, Margie Berns, Jim Miller, and Graeme Hirst, 18. Oxford: Elsevier. <https://doi.org/10.1016/B0-08-044854-2/02788-7>
- Coady, James. 1997. "L2 Vocabulary Acquisition Through Extensive Reading." In *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*, edited by James Coady and Thomas Huckin, 225–237. Cambridge: Cambridge University Press.
- Coseriu, Eugenio. 1962. *Teoría del Lenguaje y Lingüística General: Cinco Estudios*. Madrid: Gredos.
- Cowie, Anthony P. 1989. "Learners' Dictionaries – Recent Advances and Developments." In *Learners' Dictionaries: State of the Art*, edited by Makhan L. Tickoo, 42–51. Singapore: SEAMEO.
- Cowie, Anthony P. 1990. "Language as Words: Lexicography." In *An Encyclopedia of Language*, edited by N. E. Collinge, 363–377. London and New York: Routledge
- Cowie, Anthony P. 1998a. "A.S. Hornby: a Centenary Tribute." In *Proceedings of the 8th EURALEX International Congress, Liège, Belgium*, edited by Thierry Fontenelle, Philippe Hilgsmann, Archibald Michiels, André Moulin, Siegfried Theissen.
- Cowie, Anthony P. 1998b. "A.S. Hornby, 1898–1998: A Centenary Tribute." *International Journal of Lexicography* 11 (4).
- Cowie, Anthony P. 1999a. *English Dictionaries for Foreign Learners: A History*. Oxford: University Press.

- Cowie, Anthony P. 1999b. "Learner's Dictionaries in a Historical and Theoretical Perspective." In *The Perfect Learners' Dictionary* (?), edited by Thomas Herbst and Kerstin Popp, 3–14. Tübingen: Max Niemeyer Verlag. <https://doi.org/10.1515/9783110947021.3>
- Cowie, Anthony P. 2009. "The Earliest Foreign Learners' Dictionaries." In *The Oxford History of English Lexicography. Vol. 2: Specialized Dictionaries*, edited by Anthony P. Cowie, 385–411. Oxford: Clarendon Press.
- Coxhead, Averil. 2000. "A New Academic Word List." *TESOL Quarterly*, 34 (2): 213–238. <https://doi.org/10.2307/3587951>
- Crabbe, Stephen. 2017. *Controlling Language in Industry. Controlled Languages for Technical Documents*. Cham: Palgrave Macmillan. <https://doi.org/10.1007/978-3-319-52745-1>
- Croft, William, Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511803864>
- Cruse, Alan. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Cruse, Alan. 2004. *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.
- Crystal, David. 2003. *A Dictionary of Linguistics and Phonetics*. 5th ed. Oxford: Blackwell Publishing.
- Crystal, David. 1986. "The Ideal Dictionary, Lexicographer and User." In *Lexicography: An Emerging International Profession*, edited by In Robert F. Ilson, 72–81. Manchester: Manchester University Press.
- Crystal, David. 2006. "Into the Twenty-first Century." In *The Oxford History of English*, edited by Lynda Mugglestone, 394–414. Oxford: Oxford University Press.
- Cumming, Geoff, Samantha Cropp, Ronald Sussex. 1994. "On-line Lexical Resources for Language Learners: Assessment of Some Approaches to Word Definition." *System* 22(3): 369–377. [https://doi.org/10.1016/0346-251X\(94\)90022-1](https://doi.org/10.1016/0346-251X(94)90022-1)
- Cummins, Jim. 2008. "Teaching for Transfer: Challenging the Two Solitudes Assumption in Bilingual Education." *Encyclopedia of Language and Education*, 2nd edition, Vol. 5: *Bilingual Education*, edited by Jim Cummins and Nancy H. Hornberger, 65–75. New York: Springer. https://doi.org/10.1007/978-0-387-30424-3_116
- Curzan, Anne. 2000. "The Compass of the Vocabulary." In *Lexicography and the OED*, edited by Lynda Mugglestone, 96–109. Oxford: Oxford University Press.
- Dancette, Jeanne. 2003. "Les Représentations Lexico-Sémantiques (RLS), Moyen de Structuration des Connaissances dans les Domaines Spécialisés," *ISKO-France, Grenoble*, 1–16 [83–96].
- Davies, Mark. 2012. *Corpus of Contemporary American English (1990–2012)*. Accessed March 15, 2018. <http://corpus.byu.edu/coca/>
- DeRocher, James E., Murray S. Miron, S. M. Patten, Charles C. Pratt. 1973. *The Counting of Words: A Review of the History, Techniques and Theory of Word Counts with Annotated Bibliography*. New York: Syracuse University Research Corp.
- De Schryver, Gilles-Maurice. 2008. "Review of Patrick Hanks (ed.). *Lexicology: Critical Concepts*." *International Journal of Lexicography* 21 (4): 419–438. <https://doi.org/10.1093/ijl/ecno40>
- De Schryver, Gilles-Maurice, and Daniel J. Prinsloo. 2011. "Do Dictionaries Define on the Level of Their Target Users? A Case Study for Three Dutch Dictionaries." *International Journal of Lexicography* 24 (1): 5–28. <https://doi.org/10.1093/ijl/ecq045>

- Dewey, Godfrey. 1923. *Relative Frequency of English Speech Sounds*. London: Harvard University Press.
- Dodd, Stuart C. 1955. "Model English." In *Machine Translation of Languages*, edited by William N. Locke and Andrew D. Booth, 165–173. Cambridge: Technology Press of Massachusetts Institute of Technology.
- Dolan, William, Stephen Richardson, Lucy Vanderwende. 1993. "Automatically Deriving Structured Knowledge Bases from On-line Dictionaries." In *Proceedings of the First Conference of the Pacific Association for Computational Linguistics*, edited by Paul McFetridge and Fred Popowich, 5–14. Vancouver: Pacific Association for Computational Linguistics.
- Duygu, Ataman, Matteo Negri, Marco Turchi, Marcello Federico. 2017. Linguistically Motivated Vocabulary Reduction for Neural Machine Translation from Turkish to English. *The Prague Bulletin of Mathematical Linguistics* 108: 331–342. <https://doi.org/10.1515/pralin-2017-0031>
- Dziemianko, Anna. 2004. "Verb Syntax in Monolingual English Learners' Dictionaries: A Study of User-Friendliness" In Proceedings of the Eleventh EURALEX International Congress (Vol. II), Lorient, France. Université de Bretagne-Sud, edited by Geoffrey Williams and Sandra Vessier: 683–691.
- Dziemianko, Anna. 2012a. *Noun and Verb Codes in English Monolingual Dictionaries for Foreign Learners: A Study of Usefulness in the Polish Context*. Poznan: Wydawnictwo Naukowe UAM.
- Dziemianko, Anna. 2012b. "On the Use(fulness) of Paper and Electronic Dictionaries." In *Electronic Lexicography*, edited by Sylviane Granger and Magali Paquot, 319–341. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199654864.003.0015>
- Dziemianko, Anna. 2012c. "Why One and Two do not Make Three: Dictionary Form Revisited." *Lexikos* 22: 195–216. <https://doi.org/10.5788/22-1-1003>
- Dziemianko, Anna, and Robert Lew. 2006. "Research into Dictionary Use by Polish Learners of English: Some Methodological Considerations." In *IFAtuation: A Life in IFA. A Festschrift for Professor Jacek Fisiak on the Occasion of His 70th Birthday*. Poznan, edited by Katarzyna Dziubalska-Kolaczyk.
- Eder, Maciej. 2017. "Visualization in Stylometry: Cluster Analysis Using Networks." *Digital Scholarship in the Humanities* 32 (1): 50–64. <https://doi.org/10.1093/lc/fqw061>
- Eder, Maciej, Jan Rybicki, Mike Kestemont. 2016. "Stylometry with R: A Package for Computational Text Analysis." *The R Journal* 8 (1): 107–121. <https://doi.org/10.32614/RJ-2016-007>
- Ek, Jan Ate van 1975. *The Threshold Level English in a European Unit/Credit System for Modern Language Learning by Adults*. (Council of Europe Modern Languages Project). Strasbourg: Council of Europe.
- Ek, Jan Ate van, L. G. Alexander, M. A. Fitzpatrick. 1980. *Waystage English: An Intermediary Objective below Threshold Level in a European Unit/Credit System of Modern Language Learning by Adults*. Oxford: Pergamon Press.
- Ek, Jan Ate van, and John L. M. Trim. 1990a/1998a. *Threshold 1990*. Cambridge: Cambridge University Press.
- Ek, Jan Ate van, and John L. M. Trim. 1990b/1998b. *Waystage 1990*. Cambridge: Cambridge University Press.
- Ek, Jan Ate van, and John L. M. Trim. 2001. *Vantage*. Cambridge: Cambridge University Press.
- Eldridge, R. C. 1911. *Six Thousand Common English Words*. Their Comparative Frequency and What Can Be Done With Them. New York: Niagara Falls.

- Engels, L. K. 1968. *The Fallacy of Word-Counts*. IRAL, 6(3), 213–231.
- English Oxford Living Dictionaries. 2017. Accessed September 10, 2017. <https://en.oxforddictionaries.com/>
- Eunson, Baden. 1996. *Writing in Plain English*. New York: John Wiley & Sons.
- Faucett, Lawrence, and Itsu Maki. 1932. *A Study of English Word-Values, Statistically Determined from the Latest Extensive Word-Counts*. Tokyo: Matsumura Sanshodo.
- Faucett, Lawrence, Harold Palmer, Edward L. Thorndike, and Michael West. 1936. *Interim Report on Vocabulary Selection*. London: P.S. King and Son, Ltd.
- Fellbaum, Christiane, ed.. 1998. *WordNet: An Electronic Lexical Database*. Cambridge: The MIT Press. <https://doi.org/10.7551/mitpress/7287.001.0001>
- Ferraresi, Adriano, Eros Zanchetta, Marco Baroni, and Silvia Bernardini. 2008. “Introducing and Evaluating ukWaC, a Very Large Web-Derived Corpus of English.” In *Proceedings of the 4th Web as Corpus Workshop (WAC-4) Can We Beat Google? Marrakech, Morocco, 1 June 2008*, edited by Stefan Evert, Adam Kilgarrieff and Serge Sharoff. Accessed May, 4, 2017. <http://webascorpus.sourceforge.net/>
- Fischer, Ute. 1991. “How Students Misunderstand Definitions: Some Evidence for Contextual Representations of Word Meanings.” In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society, Chicago, August 1991*. Hillsdale: Lawrence Erlbaum.
- Flood, Walter E., and Michael West. 1953. “Supplementary Scientific and Technical Vocabulary.” In *A General Service List of English Words*, edited by Michael West, 583–588. London: Longmans.
- Flesch, Rudolf Franz. 1944. How Basic is Basic English? *Harper's Magazine*: 339–43.
- Fontenelle, Thierry. 1994. “What on Earth are Collocations?” *English Today* 10 (4): 42–48. <https://doi.org/10.1017/S0266078400007902>
- Fontenelle, Thierry. 2009. “Linguistic Research and Learners’ Dictionaries: the Longman Dictionary of Contemporary English.” In *Oxford History of English Lexicography*, edited by Anthony P. Cowie, 412–435. Oxford: Oxford University Press.
- Fox, Gwyneth. 1989. “A Vocabulary for Writing Dictionaries.” In *Learners’ Dictionaries: State of the Art*, edited by Makhan L. Tickoo, 153–171. Singapore: SEAMEO.
- Francis, Winthrop N., and Henry Kucera. 1964. *Manual of Information to Accompany A Standard Corpus of Present-Day Edited American English, for use with Digital Computers*. Providence, Rhode Island: Department of Linguistics, Brown University. Revised 1971. Revised and amplified 1979.
- Francis, Winthrop N., and Henry Kucera. 1982. *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin.
- Friend, Joseph Harold. 1967. *The Development of American Lexicography 1798–1864*. The Hague: Mouton.
- Fries, Charles C. and Traver, Alice A. 1963. *English Word Lists. A Study of their Adaptability for Instruction*. Ann Arbor: The George Wahr Publishing.
- Gairns, Rurth, and Redman, Stuart. 1986. *Working with Words: A Guide to Teaching and Learning Vocabulary*. Cambridge University Press.
- Gao, Jun. 2012. “Basic Cognitive Experiences and Definitions in the Longman Dictionary of Contemporary English.” *International Journal of Lexicography* 26 (1): 58–89. <https://doi.org/10.1093/ijl/ecso14>
- Gardner, Dee, and Mark Davies. 2013. “A New Academic Vocabulary List.” *Applied Linguistics* 35(3): 1–24.

- Geeraerts, Dirk. 2003. "Meaning and Definitions". In *A Practical Guide to Lexicography*, edited by Piet van Sterkenburg, 83–93. Amsterdam: John Benjamins.
<https://doi.org/10.1075/tlrp.6.11gee>
- Geeraerts, Dirk. 2006a. "The Definitional Practice of Dictionaries and the Cognitive Semantic Conception of Polysemy." *Words and other Wonders. Papers on Lexical and Semantic Topics*, edited by Dirk Geeraerts, 345–363. Berlin: Mouton de Gruyter.
<https://doi.org/10.1515/9783110219128.5.345>
- Geeraerts, Dirk. 2006b. "Prospects and Problems of Prototype Theory." In *Corpus Linguistics: Basic Readings*, edited by Dirk Geeraerts, 141–166. Berlin: Mouton de Gruyter.
<https://doi.org/10.1515/9783110199901.141>
- Geeraerts, Dirk. 2010. *Theories of Lexical Semantics*. Oxford: Oxford University Press.
- Ghadessy, Mohsen. 1979. "Frequency Counts, Word Lists and Materials Preparation: A New Approach." *English Teaching Forum* 17 (1): 24–27.
- Gill, Francis. 2015. "'Hugh's across all that' Some changing uses of prepositions." *Corpora, Grammar and Discourse In honour of Susan Hunston*, edited by Nicholas Groom, Maggie Charles, and Suganthi John, 73–96. John Benjamins Publishing Company.
<https://doi.org/10.1075/scl.73.04fra>
- Gilner, Leah. 2011. "A Primer on the General Service List." *Reading in a Foreign Language*, 23 (1): 65–83.
- Gilner, Leah, and Frank Morales. 2008. "Corpus-based Frequency Profiling: Migration to a Word List Based on the British National Corpus." *The Buckingham Journal of Language and Linguistics*: 41–58.
- Goddard, Cliff. 2010. "The Natural Semantic Metalanguage Approach." *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press: 459–484.
- Goddard, Cliff, ed.. 2018. *Minimal English for a Global World. Improved Communication Using Fewer Words*. Cham: Palgrave Macmillan. <https://doi.org/10.1007/978-3-319-62512-6>
- Goddard, Cliff, and Anna Wierzbicka, eds.. 2002. *Meaning and Universal Grammar – Theory and Empirical Findings*. Vols I and II. John Benjamins.
<https://doi.org/10.1075/slcs.6010.1075/slcs.61>
- Goddard, Cliff, and Anna Wierzbicka, eds.. 2014. *Words and Meanings: Lexical Semantics across Domains, Languages and Cultures*. Oxford: Oxford University Press.
- Goddard, Cliff, and Anna Wierzbicka. 2018. "Minimal English and How it can Add to Global English." In *Minimal English for a Global World. Improved Communication Using Fewer Words*, edited by Cliff Goddard, 5–28. Cham: Palgrave Macmillan.
https://doi.org/10.1007/978-3-319-62512-6_2
- Gordon, W. Terrence, ed.. 1994. *C.K. Ogden and Linguistics*. 5 volumes. London.
- Gove, Phillip B. 1972. "English Dictionaries of the Future." In *New Aspects of Lexicography*, edited by Howard D. Weinbrot, 152–170. London: Southern Illinois University Press.
- Grabowski, Łukasz. 2015. *Phraseology in English Pharmaceutical Discourse. A Corpus-Driven Study of Register Variation*. Opole: Wydawnictwo Uniwersytetu Opolskiego.
- Greenbaum, Sidney, Charles F. Meyer, and John Taylor. 1984. "The Image of the Dictionary for American College Students." *Dictionaries* 6: 31–52. <https://doi.org/10.1353/dic.1984.0008>
- Grochowski, Maciej. 1993. *Konwencje Semantyczne a Definiowanie Wyrażeń Językowych*. Warszawa: Zakład Semiotyki Logicznej Uniwersytetu Warszawskiego.
- Gries, Stefan Th. 2009. *Quantitative Corpus Linguistics with R*. Routledge: New York.

- Grzegorzczkova, Renata. 2001. *Wprowadzenie do Semantyki Językoznawczej*. Warszawa: Wydawnictwo Naukowe PWN.
- Hancioglu, Nilgün, Steven Neufeld, and John Eldridge. 2008. "Through the Looking Glass and into the Land of Lexico-Grammar." *English for Specific Purposes* 27: 459–79. <https://doi.org/10.1016/j.esp.2008.08.001>
- Hanks, Patrick. 1979. "To What Extent does a Dictionary Definition Define?" In *Dictionaries and their Users*, edited by Reinhard R. K. Hartmann. Exeter: University of Exeter, 32–38.
- Hanks, Patrick. 1987. "Definitions and Explanations." In *Looking up: An account of the COBUILD project in lexical computing*, edited by John M. Sinclair, 116–136. London, Glasgow.
- Hanks, Patrick. 1990. 'Evidence and Intuition in Lexicography', In *Meaning and Lexicography*, edited by Jerzy Tomaszczyk and Barbara Lewandowska-Tomaszczyk, 31–41. Amsterdam and Philadelphia: John Benjamins. <https://doi.org/10.1075/lsee.28.07han>
- Hanks, Patrick. 1994. "Lexicography: Theory and Practice." *Dictionaries*. 1992/3 (14): 97–112.
- Hanks, Patrick. 2005. "Johnson and Modern Lexicography." *International Journal of Lexicography* 18 (2): 243–266. <https://doi.org/10.1093/ijl/ecio24>
- Hanks, Patrick. 2008. *Lexicology: Critical Concepts*. Abingdon: Routledge.
- Hanks, Patrick. 2009. Common Sense Blossoms in Springfield, MA. Review of Merriam-Webster's Advanced Learner's English Dictionary. Springfield, MA: Merriam-Webster Inc, edited by Stephen J. Perrault, *International Journal of Lexicography* 22 (3).
- Hanks, Patrick. 2015. Cognitive Semantics and the Lexicon. Review Article. *International Journal of Lexicography* 28 (1): 86–106. <https://doi.org/10.1093/ijl/ecv002>
- Hanks, Patrick. 2019. *Pattern Dictionary of English Verbs (PDEV)*. <https://pdev.org.uk/>
- Hanks, Patrick, and James Pustejovsky. 2005. "A Pattern Dictionary for Natural Language Processing." *Revue Française de Linguistique Appliquée*, 10 (2): 63–82. <https://doi.org/10.3917/rfla.102.82>
- Hartmann, Reinhard R. K. 1989. "Dictionaries of English: The User's Perspective." In *Dictionaries. Prospects for the Record of Our Language*, edited by Richard W. Bailey, 121–135. Cambridge: Cambridge University Press.
- Hartmann, Reinhard R. K. 1993. "Recent trends in pedagogical lexicography: The case of the 'bilingualised' learner's dictionary". In *Endangered Languages. Proceedings of XVth International Congress of Linguists, Québec, August 1992*, edited by André Crochetière, Jean-Claude Boulanger, and Conrad Ouellette, 159–162. Québec: Les Presses de l'Université Laval.
- Hartmann, Reinhard R. K., and Gregory James. 1998. *Dictionary of Lexicography*. London: Routledge. <https://doi.org/10.4324/9780203159040>
- Harvey, Keith, and Deborah Yuill. 1997. "A Study of the Use of a Monolingual Pedagogical Dictionary by Learners of English Engaged in Writing." *Applied Linguistics* 18 (3): 253–278. <https://doi.org/10.1093/applin/18.3.253>
- Heatley, Alex, Paul I. S. Nation, and Averil Coxhead. 2002. *RANGE and FREQUENCY Programs*. <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>.
- Herbst, Thomas. 1986. "Defining with a Controlled Defining Vocabulary in Foreign Learners' Dictionaries." *Lexicographica: International Annual for Lexicography* 2: 101–119.
- Herbst, Thomas. 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. <https://doi.org/10.1093/ijl/9.4.321>
- Hiles, Lorna. 2014. "Towards a Southern African English Defining Vocabulary." *Lexikos* 24: 178–185. <https://doi.org/10.5788/24-1-1257>

- Hindmarsh, Roland. 1980. *The Cambridge English Lexicon*. Cambridge: Cambridge University Press.
- Hirsh, David. 1993. *The vocabulary demands and vocabulary learning opportunities in short novels*. Unpublished master's thesis, Victoria University of Wellington, New Zealand.
- Hirsh, David, and Paul I. S. Nation. 1992. "What Vocabulary Size is Needed to Read Unsimplified Texts for Pleasure?" *Reading in a Foreign Language*, 8 (2): 689–696.
- Horn, Ernest. 1926. *A Basic Writing Vocabulary*. Iowa City: State University of Iowa Monograph First Series No. 4.
- Hornby, Albert S. 1938. "Report on the Year's Work: Dictionary Problems." *IRET Bulletin* 148: 20–8; 36.
- Horst, Marlise, Tom Cobb, and Paul Meara. 1998. "Beyond a Clockwork Orange: Acquiring Second Language Vocabulary through Reading." *Reading in a Foreign Language* 11: 207–223.
- Howatt, Anthony Phillip R. 1984. *A History of English Language Teaching*. Oxford: Oxford University Press.
- Howatt, Anthony Phillip R. 2006. "Language Teaching: History." In *Encyclopedia of Language and Linguistics*, 2nd ed., edited by Keith Brown, Anne Anderson, Laurie Bauer, Margie Berns, Jim Miller, and Graeme Hirst, 634–646. Oxford: Elsevier.
<https://doi.org/10.1016/B0-08-044854-2/01325-0>
- Hu, Marcella and Paul Nation. 2000. "Vocabulary Density and Reading Comprehension" *Reading in a Foreign Language*, 13 (1): 403–430.
- Huijsen, Willem. 1998. "Controlled Language – An Introduction." In *Proceedings of the Second International Workshop on Controlled Language Applications*. (CLAW) Pittsburgh: Carnegie Mellon University Language Technologies Institute. 1–15.
- Hulstijn, Jan H. 1993. "When do Foreign-Language Readers Look Up the Meaning of Unfamiliar Words? The Influence of Task and Learner Variables." *The Modern Language Journal* 77: 139–147. <https://doi.org/10.1111/j.1540-4781.1993.tb01957.x>
- Hutchins, W. John. 1995. "Machine Translation: A Brief History." In *Concise History of the Language Sciences: From the Sumerians to the Cognitivists*, edited by E. F. K. Koerner and Ronald E. Asher, 431–445. New York: Pergamon. <https://doi.org/10.1016/B978-0-08-042580-1.50066-0>
- Hwang, Kyongho. 1989. *Reading Newspapers for the Improvement of Vocabulary and Reading Skills*. Unpublished M.A. Thesis, Victoria University, Wellington.
- Ilson, Robert. 1986. "Lexicographic Archeology: Comparing Dictionaries of the Same Family." In *The History of Lexicography*, edited by Reinhard R. K. Hartmann, 127–136. Amsterdam: John Benjamins. <https://doi.org/10.1075/sihols.40.14ils>
- Ilson, Robert. 1990. "Present-Day British Lexicography." In *Wörterbücher Dictionaries Dictionnaires. An International Encyclopedia of Lexicography*. Vol. 2 (1989–91), edited by Franz J. Hausmann, Oskar Reichmann, Herbert E. Wiegand, and Ladislav Zgusta, 1967–1983. Berlin: Walter de Gruyter.
- Jain, Mahavir P. 1981. "On Meaning in the Foreign Learner's Dictionary." *Applied Linguistics* 2 (3): 274–286. <https://doi.org/10.1093/applin/2.3.274>
- James, Gregory C. A. 2003. "Bilingualisation as a Genre." In *Lexicography. Critical Concepts*. Vol. III, edited by Reinhard R. K. Hartmann, 135–146. London: Routledge.
- Jansen, Jacques, Mergeai, Jean-Paul, and Jocelyne Vanandroye. 1987. "Controlling LDOCE's controlled vocabulary." In *The Dictionary and the Language Learner* (Lexicographica, Series maior 17), edited by Anthony P. Cowie. Tübingen: Max Niemeyer.
<https://doi.org/10.1515/9783111340500-008>

- Jeffery, G. B. 1953. "Foreword" In *A General Service List of English Words*, edited by Michael West, v–vi. London: Longmans.
- Jenkins, Joseph R., Marcy L. Stein, and Katherine Wysocki. 1984. "Learning Vocabulary through Reading." *American Educational Research Journal* 21: 767–787.
<https://doi.org/10.3102/00028312021004767>
- Johansson, Stig. 1978. *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English for Use with Digital Computers*. Oslo: University of Oslo, Department of English.
- Johnson, Samuel. 1747. *The Plan of an English Dictionary*. London.
- Kamińska, Małgorzata Anna. 2014. *A History of the Concise Oxford Dictionary*. Frankfurt am Main: Peter Lang. <https://doi.org/10.3726/978-3-653-04370-9>
- Kamiński, Mariusz Piotr. 2013. *A History of the Chambers Dictionary*. Lexicographica Series Maior 143. Berlin: Walter De Gruyter. <https://doi.org/10.1515/9783110312737>
- Kamiński, Mariusz Piotr. 2016. "Corpus-based Extraction of Collocations for Near-Synonym Discrimination." In *Proceedings of the XVII EURALEX International Congress. Lexicography and Linguistic Diversity. 6–10 September 2016 Tbilisi*, edited by Tinatin Margalitadze and George Meladze, 375–382. <http://euralex2016.tsu.ge/publication.html>
- Kamiński, Mariusz Piotr. 2019. "Use of Hedges in Definitions: Out of Necessity or Theory Driven?" *Lexikos* 29: 55–74. <https://doi.org/10.5788/29-1-1512>
- Kamiński, Mariusz Piotr. 2020. Does a limited defining vocabulary make definitions syntactically more complex? *Studia Anglica Posnaniensia* 55: 5–19.
- Kamprath C, Adolphson E, Mitamura, T and Nyberg E 1998. "Controlled Language for Multilingual Document Production: Experience with Caterpillar Technical English." In *Proceedings of the Second International Workshop on Controlled Language Applications*, 51–61. Pittsburgh: Carnegie Mellon University Language Technologies Institute.
- Kay, Christian J. 2000. "Historical Semantics and Historical Lexicography: Will the Twain Ever Meet?" In *Lexicology, Semantics and Lexicography*, edited by Julie Coleman and Christain J. Kay, 53–68. Amsterdam: John Benjamins. <https://doi.org/10.1075/cilt.194.07kay>
- Kernerman, Lionel. 1996. "English Learners' Dictionaries: How Much Do We Know about their Use?" In *EURALEX' 96 Proceedings*. Göteborg: Department of Swedish, Göteborg University, edited by Martin Gellerstam, Jerker Jarborg, Sven-Göran Malmgren, Lena Rogström, Catarina Røjder Papmehl.
- Kernerman, Ilan. 1999. "Back to the Future." *Dictionary News*, no 7: 1.
- Khoong, Elaine, C., Eric Steinbrook, Cortlyn Brown, and Alicia Fernandez. 2019. "Assessing the Use of Google Translate for Spanish and Chinese Translations of Emergency Department Discharge Instructions." *JAMA Internal Medicine* 2019, E1–E3.
- Kilgarrieff, Adam. 2006. "Word Senses." In *Word Sense Disambiguation: Algorithms and Applications*, edited by Eneko Agirre and Phillip Edmonds, 29–45. New York: Springer.
https://doi.org/10.1007/978-1-4020-4809-8_2
- Kilgarrieff, Adam, and Michael Rundell. 2002. "Lexical Profiling Software and its Lexicographic Applications: Case Study." In *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002*, Copenhagen: Center for Sprogteknologi, edited by Anna Braasch and Claus Povlsen, 807–819.
- Kilgarrieff, Adam, Joseph Rosenzweig. 1999. "English SENSEVAL: reports and results." In *Proceedings of the 5th natural language processing Pacific Rim symposium (NLPRS '99)*, Beijing, China.

- Kirkness, Alan. 2001. "Review. English Dictionaries for Foreign Learners: A History. A. P. Cowie." *ELT Journal* 55 (1): 110–112. <https://doi.org/10.1093/elt/55.1.110>
- Kirkpatrick, Betty. 1985. "A Lexicographical Dilemma: Monolingual Dictionaries for the Native Speaker and for the Learner." In *Dictionaries, Lexicography and Language Learning*, edited by Robert Ilson, 7–13. Oxford 1985 (ELT Documents 120).
- Kitzes, Adam H. 2006. *The Politics of Melancholy from Spenser to Milton*. New York: Routledge.
- Kjellmer, Göran. 1986. "'The Lesser Man': Observations on the Role of Women in Modern English Writings." In *Corpus Linguistics II*, edited by Jan Arts and Willem Meijs, 163–76. Amsterdam: Rodopi.
- Klare, George R. 1968. "The Role of Word Frequency in Readability." *Elementary English* 45 (1): 12–22.
- Knowles, J. 1904. *The London Point System of Reading for the Blind*. London.
- Koeneke, Rodney. 2004. *Empires of the Mind. I. A. Richards and Basic English in China, 1929–1979*. Stanford: Stanford University Press.
- Krashen, Stephen D. 2011. *Free Voluntary Reading*. Santa Barbara: Libraries Unlimited.
- Kucera, Henry, and Winthrop N. Francis. 1967. *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.
- Landau, Sidney I. 1993. "Wierzbicka's Theory and the Practice of Lexicography." *Dictionaries*. 14: 113–119. <https://doi.org/10.1353/dic.1992.0007>
- Landau, Sidney I. 2001. *Dictionaries: The Art and Craft of Lexicography*. 2nd edition. Cambridge: University Press.
- Landau, Sidney I. 2009. "Major American Dictionaries." In *The Oxford History of English Lexicography*. Vol. 2, edited by Anthony P. Cowie, 182–229. Oxford: Oxford University Press.
- Laufer, Batia. 1989. "What Percentage of Text Lexis is Essential for Comprehension?" In *Special Language: From Humans Thinking to Thinking Machines*, edited by Christer Lauren and Marianne Nordman, 316–323. Clevedon: Multilingual Matters.
- Laufer, Batia. 1993. "The Effect of Dictionary Definitions and Examples on the Use and Comprehension of New L2 Words." *Cahiers de Lexicologie* 63 (2): 131–142.
- Laufer, Batia, Cathie Elder, Kathryn Hill, and Peter Congdon. 2004. "Size and strength: Do we need both to measure vocabulary knowledge?" *Language Testing* 21: 202–226. <https://doi.org/10.1191/0265532204lt2770a>
- Leech, Geoffrey, Paul Rayson, and Andrew Wilson. 2001. *Word Frequencies in Written and Spoken English*. Harlow: Pearson Education Limited.
- Lew, Robert. 2002. "A Study in the Use of Bilingual and Monolingual Dictionaries by Polish Learners of English: A Preliminary Report." In *Proceedings of the Tenth EURALEX International Congress (Vol. II)*, Copenhagen: Center for Sprogteknologi, edited by A. Anna Braasch and Claus Povlsen, 759–763.
- Lew, Robert. 2004. *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-Bilingual Dictionaries by Polish Learners of English*. Poznan: Motiwex.
- Lew, Robert. 2010. "Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries." *Lexikos* 20: 290–306. <https://doi.org/10.4314/lex.v20i1.62717>
- Lew, Robert. 2012. "How Can We Make Electronic Dictionaries More Effective?" In *Electronic Lexicography*, edited by Sylviane Granger and Magali Paquot, 343–363. Oxford: Oxford University Press.
- Lew, Robert. 2013. "Identifying, Ordering and Defining Senses." In *The Bloomsbury Companion to Lexicography*, edited by Howard Jackson, 284–302. London: Bloomsbury.

- Lew, Robert, Anna Dziemianko. 2006. "A New Type of Folk-inspired Definition in English Monolingual Learners' Dictionaries and its Usefulness for Conveying Syntactic Information." *International Journal of Lexicography* 19 (3): 225–242.
<https://doi.org/10.1093/ijl/eclo11>
- Lewandowska-Tomaszczyk, Barbara. 2007. "Polysemy, Prototypes, and Radial Categories." In *The Oxford Handbook of Cognitive Linguistics*, edited by Dirk Geeraerts and Hubert Cuyckens, 139–169. Oxford: Oxford University Press.
- Li, Hanhong. 2010. "Word Frequency Distribution for Electronic Learner's Dictionaries." In *eLexicography in the 21st Century: New Challenges, New Applications. Proceedings of eLex 2009, Louvain-la-Neuve, 22–24 October 2009*, edited by Sylviane Granger, Magali Paquot, 203–216. UCL Presses Universitaires de Louvain.
- Lorge, Irving. 1944. "Word Lists as Background for Communication." *Teachers College Record* 45: 543–552.
- Lorge, Irving. 1949. *Semantic Count of the 570 Commonest English Words*. New York: Teachers College, Columbia University.
- Lorge, Irving. 1953. "The Semantic Count." In *A General Service List of English Words*, edited by Michael West, xi–xiii. London: Longmans.
- Lorge, Irving and Edward L. Thorndike. 1938. *A Semantic Count of English Words*. New York: Teachers College, Columbia University.
- Louw, Phillip. 2000. "An Integrated Semasiological and Onomasiological Presentation of Semantic Information in General Monolingual Dictionaries as Proposed in H.E. Wiegand's Semantics and Lexicography." *Lexikos* 10: 119–137.
- Lovelace, Eugene A. 1988. "On Using Norms for Low-Frequency Words." *Bulletin of the Psychonomic Society* 26 (5): 410–412. <https://doi.org/10.3758/BF03334899>
- Lupescu, Stuart, and Richard Day. 1993. "Reading, Dictionaries and Vocabulary Learning." *Language Learning* 43 (1): 263–287. <https://doi.org/10.1111/j.1467-1770.1992.tb00717.x>
- Lynn, Robert W. 1973. "Preparing Word Lists: A Suggested Method." *RELC Journal* 11 (2): 25–32. <https://doi.org/10.1177/003368827300400103>
- Lyons, John. 1969. *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- Lyons, John. 1977. *Semantics*. 2 vols. Cambridge: Cambridge University Press.
- Maat, Jaap. 2004. *Philosophical Languages in the Seventeenth Century: Dalgarno, Wilkins, Leibniz*. Dordrecht: Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-007-1036-8>
- MacFarquhar, Peter. 1985. *ESL Learner Dictionaries: A Study of Definition Difficulty*. Occasional Paper 8. University of Hawaii at Manoa.
- MacFarquhar, Peter D., and Jack C. Richards. 1983. "On Dictionaries and Definitions." *RELC Journal* 14 (1): 111–24. <https://doi.org/10.1177/003368828301400112>
- Martinez, Maria-Angeles. 2015. "Modality Awareness Development in Narratives for Young Readers of English." In *Thinking Modally: English and Contrastive Studies on Modality*, edited by Juan Rafael Zamorano-Mansilla, Carmen Maiz, Elena Dominguez, Victoria Martin De la Rosa, 429–448. Cambridge: Cambridge Scholars Publishing.
- McArthur, Tom. 1986. *Worlds of Reference*. Cambridge: University Press.
- McArthur, Tom. 1998a. *Living Words. Language, Lexicography and the Knowledge Revolution*. Exeter: University of Exeter Press.
- McArthur, Tom. 1998b. "The Background and Nature of ELT Learner's Dictionaries." In *Learners' Dictionaries: State of the Art*, edited by Makhan L. Tickoo, 52–64. Singapore: SEAMEO.

- McConchie, Rod W. 2013. "The Most Discriminating Plagiarist': The Unkindest Cut (and Paste) of All". In *Selected Proceedings of the 2012 Symposium on New Approaches in English Historical Lexis (HEL-LEX 3)*, edited by Rod W. McConchie, Teo Juvonen, Mark Kaunisto, 107–119. Somerville, MA: Cascadilla Proceedings Project.
- McCreary, Don R., and Elise Amacker. 2006. "Experimental Research on College Students' Usage of Two Dictionaries: A Comparison of the Merriam-Webster College Dictionary and the Macmillan English Dictionary for Advanced Learners". In *Proceedings of 12th EURALEX International Congress, EURALEX 2006*, edited by Elisa Corino, Carla Marello, and Cristina Onesti, 871–885. Alessandria: Edizioni Dell'Orso.
- McCreary, Don R., and Fredric T. Dolezal. 1999. "A Study of Dictionary Use by ESL Students in an American University." *International Journal of Lexicography* 12 (2): 107–146. <https://doi.org/10.1093/ijl/12.2.107>
- McKeown, Margaret. 1993. "Creating Effective Definitions for Young Word Learners." *Reading Research Quarterly* 28 (1): 16–31. <https://doi.org/10.2307/747814>
- Meer, Geart van der. 1999. "Metaphors and Dictionaries: The Morass of Meaning, or How to Get Two Ideas for One." *International Journal of Lexicography* 12 (3): 195–208. <https://doi.org/10.1093/ijl/12.3.195>
- Meer, Geart van der. 2000. "Core, Subsense, and the New Oxford Dictionary of English (NODE)." In *Proceedings of the Ninth EURALEX Congress*, edited by Ulrich Heid, Stefan Evert, Egbert Lehmann, Christian Rohrer, 419–431. Stuttgart: Institut für Maschinelle Sprachverarbeitung.
- Meer, Geart van der. 2004. "On Defining: Polysemy, Core Meanings, and 'Great Simplicity.'" In *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*, Vol. 2, edited by Geoffrey Williams and Sandra Vessier, 807–15. Lorient: Université de Bretagne Sud.
- Michiels, Archibald, and Jacques Noël. 1984. "The Pro's and Con's of a Controlled Defining Vocabulary in a Learner's Dictionary." In *LEXeter's 83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*, edited by Reinhard R. K. Hartmann, 385–394. Tübingen: Max Niemeyer Verlag. <https://doi.org/10.1515/9783111593166-051>
- Miller, George A., and Patricia M. Gildea. 1987. "How Children Learn Words." *Scientific American* 257 (3): 94–99. <https://doi.org/10.1038/scientificamericano987-94>
- Minkova, Donka, and Robert Stockwell. 2006. "English Words." In *The Handbook of English Linguistics*, edited by Bas Aarts and April McMahon. London: Blackwell Publishing. 461–482. <https://doi.org/10.1002/9780470753002.ch20>
- Miodunka, Wladyslaw. 1989. *Podstawy Leksykologii i Leksykografii*. Warsaw: Panstwowe Wydawnictwo Naukowe.
- Mitamura, Teruko. 1999. "Controlled Language for Multilingual Machine Translation." In *Machine Translation Summit VII, Singapore, 1999*, 46–52.
- Moon, Rosamund. 1987. "The Analysis of Meaning." In *Looking up: The COBUILD Project in Lexical Computing*, edited by John Sinclair. Collins ELT: London and Glasgow.
- Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English*. Oxford: Clarendon Press.
- Morton, Herbert C. 1994. *The Story of Webster's Third. Philip Gove's Controversial Dictionary and its Critics*. Cambridge: Cambridge University Press.
- Muegge, Uwe. 2008. *Machine Translation. Controlled Language. Translation Standards*. Accessed August 1, 2017. <http://www.muegge.cc/controlled-language.htm>

- Murray, James A. H. 1933. "General Explanations." In *The Oxford English Dictionary, Being a Corrected Re-issue with an Introduction, Supplement, and Bibliography of A New English Dictionary on Historical Principles*, edited by James A. H. Murray et al. Oxford: The Clarendon Press.
- Nakane, Sadayuki. 1998. "The Defining Vocabularies of LDOCE, OALD, and CIDE: A Comparative Study." *Bulletin of Liberal Arts* 18.
- Nation, I. S. Paul. 1990. *Teaching and Learning Vocabulary*. New York: Newbury House.
- Nation, I. S. Paul. 1993. "Measuring Readiness for Simplified Material: A Test of the First 1,000 Words of English." In *Simplification: Theory and Application. RELC Anthology Series 31*, edited by Makhan L. Tickoo, 193–203.
- Nation, I. S. Paul. 2004. "A Study of the Most Frequent Word Families in the British National Corpus." In *Vocabulary in a Second Language*, edited by Paul Bogaards and Batia Laufer, 3–13. Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.10.03nat>
- Nation, Paul. 2006. "How Large a Vocabulary is Needed for Reading and Listening?." *Canadian Modern Language Review* 63 (1), 59–82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, Paul. 2009. "New Roles for L2 Vocabulary?." In *Contemporary Applied Linguistics. Volume One. Language Teaching and Learning*, edited by Vivian Cook and Li Wei, 99–116.
- Nation, Paul. 2012. "Information on the BNC/COCA lists." Accessed August 16, 2018. <https://www.victoria.ac.nz/lals/about/staff/paul-nation#vocab-lists>
- Nation, I. S. Paul, and Alex Heatley. 2002. *Range: A Program for the Analysis of Vocabulary in Texts [software]*. <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>
- Nation, I. S. Paul, and Kyongho Hwang. 1995. "Where Would General Service Vocabulary Stop and Special Purposes Vocabulary Begin?" *System* 23 (1): 35–41. [https://doi.org/10.1016/0346-251X\(94\)00050-G](https://doi.org/10.1016/0346-251X(94)00050-G)
- Nation, I. S. P. and Robert Waring. 1997. "Vocabulary Size, Text Coverage and Word Lists." In *Vocabulary: Description, Acquisition and Pedagogy*, edited by Norbert Schmitt and Michael McCarthy, 6–19. Cambridge: Cambridge University Press. Accessed February 15, 2018. <http://www.robwaring.org/papers/CUP/cup.html>
- Nation, I. S. Paul, and Stuart Webb. 2011. *Researching and Analyzing Vocabulary*. Boston: Heinle.
- Nerlich, Brigitte. 1992. *Semantic Theories in Europe 1830–1930*. Amsterdam: John Benjamins. <https://doi.org/10.1075/sihols.59>
- Nesi, Hilary. 1998. "Defining a Shoehorn: The Success of Learners' Dictionary Entries for Concrete Nouns." In *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*, edited by Beryl T. Sue Atkins, 159–178. Tübingen: Max Niemeyer Verlag.
- Nesi, Hilary. 2000. *The Use and Abuse of Learners' Dictionaries*. Lexicographica. Series Maior 98. Tübingen: Max Niemeyer. <https://doi.org/10.1515/9783110946031>
- Nesi, Hilary. 2009. "Dictionaries in Electronic Form." In *The Oxford History of English Lexicography*. Vol. 2, edited by Anthony P. Cowie. Oxford: Oxford University Press.
- Nesi, Hilary. 2014. "Dictionary Use by English Language Learners." *Language Teaching* 47 (1): 38–55. <https://doi.org/10.1017/S0261444813000402>
- Nesi, Hilary, and Richard Haill. 2002. "A Study of Dictionary Use by International Students at a British University." *International Journal of Lexicography* 15 (4): 277–305. <https://doi.org/10.1093/ijl/15.4.277>
- Nesi, Hilary, and Paul Meara. 1994. "Patterns of Misrepresentation in the Productive Use of EFL Dictionary Definitions." *System* 22 (1): 1–15. [https://doi.org/10.1016/0346-251X\(94\)90036-1](https://doi.org/10.1016/0346-251X(94)90036-1)
- Nesselhauf, Nadja. 2005. *Collocations in a Learner Corpus*. Amsterdam: Benjamins. <https://doi.org/10.1075/scl.14>

- Neubauer, Fritz. 1980. *Die Struktur der Explikationen in Deutschen Einsprachigen Wörterbüchern. Eine Vergleichende Lexiko-Semantische Analyse*. Hamburg.
- Neubauer, Fritz. 1984. "The Language of Explanation in Monolingual Dictionaries," *LEXeter '83. Proceedings. Papers from the International Conference on Lexicography at Exeter*, 9–12 September 1983. *Lexicographica. Series Maior* 1, edited by Reinhard R. K. Hartmann, 117–123. Tübingen: Max Niemeyer. <https://doi.org/10.1515/978311593166-017>
- Neubauer, Fritz. 1987. "How to Define a Defining Vocabulary." *A Spectrum of Lexicography. Papers from AILA Brussels 1984*, edited by Robert Ilson, 49–60. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.28.05neu>
- Neubauer, Fritz. 1989. "Vocabulary Control in the Definitions and Examples of Monolingual Dictionaries." In *Wörterbücher Dictionnaires Dictionnaires. An International Encyclopedia of Lexicography*. 3 vols. (1989–91), edited by Franz J. Hausmann, Oskar Reichmann, Herbert E. Wiegand, and Ladislav Zgusta, 899–904. Berlin: Walter de Gruyter.
- Neufeld, Steve, and Ali Billuroglu. 2005. "In Search of the Critical Lexical Mass: How 'General' in the GSL? How 'Academic' is the AWL?" Accessed March 16, 2018. https://www.researchgate.net/publication/317661877_In_search_of_the_critical_lexical_mass_How_%27general%27in_the_GSL_How_%27academic%27is_the_AWL
- Nichols, Wendalyn. 2001. Review of A.P. Cowie. *English Dictionaries for Foreign learners: A History*. *Dictionaries* 22: 205–210.
- Nida, Eugene A. 1975. *Componential Analysis of Meaning*. The Hague: Mouton.
- Nyberg, Eric H., Teruko Mitamura, and Willem-Olaf Huijsen. 2003. "Controlled Language for Authoring and Translation." In *Computers and Translation: A Translator's Guide*, edited by Harold L. Somers, 245–281. Amsterdam and Philadelphia: John Benjamins. <https://doi.org/10.1075/btl.35.17nyb>
- Nyberg, Eric H., and Teruko Mitamura. 1992. "The KANT System: Fast, Accurate, High-Quality Translation in Practical Domains." In *Proceedings of COLLING-92, Nantes, France, July 1992, Vol. 3*, 1069–1073.
- Oakes, Michael P. 2009. "Corpus Linguistics and Language Variation." In *Contemporary Corpus Linguistics*, edited by Paul Baker, 159–83. London: Continuum.
- Odlin, Terence. 1989. *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524537>
- Ogden, Charles K. 1930. *Basic English: A General Introduction with Rules and Grammar*. London: Kegan Paul, Trench, Trubner & Co., Ltd.
- Ogden, Charles K. 1931. *Debabelization. With a Survey of Contemporary Opinion on the Problem of a Universal Language*. London: Kegan Paul, Trench, Trubner & Co., Ltd.
- Ogden, Charles K., and Ivor A. Richards. 1923. *The Meaning of Meaning*. London.
- Ogden, Charles K. 1932. *The Basic Dictionary*. London.
- Ogden, Charles K. 1940. *The General Basic English Dictionary*. London: Evans.
- Osgood, Charles E., George J. Suci, and Percy H. Tannenbaum. 1957. *The Measurement of Meaning*. Urbana: University of Illinois Press.
- Page, Edward C. 2001. *Governing by Numbers Delegated Legislation and Everyday Policy-Making*. Portland: Hart Publishing.
- Palmer, Harold E. 1929a. "What shall We Call 'a Word'?" *IRET Bulletin* 54: 1–2.
- Palmer, Harold E. 1929b. "Editorial: What is an idiom?" *IRET Bulletin* 56: 1–2.
- Palmer, Harold E. 1932. "Some Notes on Construction-Patterns" *Institute Leaflet* 38. Tokyo: IRET.

- Palmer, Harol E. 1933. *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.
- Palmer, Harol E. 1936. "The History and Present State of the Movement Towards Vocabulary Control." *IRET Bulletin* 120: 14–17; 121: 19–23.
- Palmer, Harol, and Albert S. Hornby. 1937. *Thousand-Word English*. London: George Harrap.
- Piotrowski, Tadeusz. 1994. *Problems in Bilingual Lexicography*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Piotrowski, Tadeusz. 2001. *Zrozumieć leksykografię*. Warszawa: Wydawnictwo Naukowe PWN.
- Piotrowski, Tadeusz. 2002. *An Overview of Monolingual EFL Dictionaries Today*. Accessed May 5, 2017. <http://www.academia.edu/23933212/>
- Piotrowski, Tadeusz. 2009. "Pedagogical Dictionaries of English as Commercial Products." In *On Language Structure, Acquisition and Teaching. Studies in Honour of Janusz Arabski on the Occasion of his 70th Birthday*, edited by Maria Wysocka, 171–176. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Plain English Campaign*. Accessed March 2, 2019. <http://www.plainenglish.co.uk/about-us.html>
- Plain Speaking is no Joke. 1997. *BBC News*. Accessed March 2, 2019. http://news.bbc.co.uk/2/hi/uk_news/38503.stm
- Podhajecka, Mirosława. 2009. "Historical Development of Lexicographical Genres: Some Methodological Issues." In *Selected Proceedings of the 2008 Symposium on New Approaches in English Historical Lexis (HEL-LEX 2)*, edited by Roderick W. McConchie, Alpo Honkapohja, and Jukka Tyrkkö, 153–170. Somerville, MA: Cascadilla Proceedings Project.
- Podhajecka, Mirosława. 2016. *A History of Polish-English / English-Polish Bilingual Lexicography (1788–1947)*. Opole: Wydawnictwo Uniwersytetu Opolskiego.
- Pottier, Bernard. 1964. Vers une Sémantique Moderne. *Travaux de Linguistique et de Littérature* 2: 107–137.
- Praninskas, Jean. 1972. *American University Word List*. London: Longman.
- Preiss, Judita, and Mark Stevenson. 2004. "Introduction to the Special Issue on Word Sense Disambiguation." In *Computer Speech and Language* 18 (3): 201–207. <https://doi.org/10.1016/j.csl.2004.05.005>
- Pym, Peter J. 1990. "Pre-Editing and the Use of Simplified Writing for MT: An Engineer's Experience of Operating an MT System." In *Translating and the Computer 10: The Translation Environment 10 Years on*, edited by Pamela Mayorcás, 80–95. London: Aslib.
- Quirk, Randolph. 1973. "The Social Impact of Dictionaries in the U.K." In *Lexicography in English*, edited by Raven I. McDavid and Audrey Duckert, 76–88. New York: New York Academy of Sciences.
- Quirk, Randolph. 1982. "International communication and the concept of nuclear English." In *English for Cross-cultural Communication*, edited by Larry E. Smith, 151–165. New York: St. Martin's Press. [Re-printed in *Style and Communication in the English Language*. London: Arnold, 37–53.]
- Quirk, Randolph. 1986. "Opening Remarks." In *Lexicography. An Emerging International Profession*, edited by Robert Ilson, 1–6. Manchester: Manchester University Press.
- Qun, Liu, and Zhang Xiaojun. 2015. "Machine Translation." In *The Routledge Encyclopedia of Translation Technology*, edited by Chan Sin-wai, 105–119. Routledge: New York.
- R Development Core Team. 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Read, John. 1988. "Measuring the vocabulary knowledge of second language learners." *RELC Journal* 19 (2): 12–25. <https://doi.org/10.1177/003368828801900202>

- Redmond, Phil. 2010. "Chrissie Maher—a story of inspiration and aspiration." *Liverpool Daily Post*. Liverpool: Trinity Mirror North West & North Wales.
- Reisner, Edward H. 1925. *Nationalism and Education since 1789: a Social and Political History of Modern Education*. London: The Macmillan Company.
- Renouf, Antoinette. 1987. "Corpus Development" In *Looking Up: An Account of the COBUILD Project in Lexical Computing*, edited by John Sinclair, 10–40. London: HarperCollins.
- Rey, Alain. 1990. "Definitional Semantics: Its Evolution in French Lexicography." *Meaning and Lexicography*, edited by Jerzy Tomaszczyk and Barbara Lewandowska-Tomaszczyk, 43–55. Amsterdam: John Benjamins. <https://doi.org/10.1075/llsee.28.08rey>
- Riemer, Nick. 2006. "Reductive Paraphrase and Meaning: A Critique of Wierzbickian Semantics." *Linguistics and Philosophy* 29: 347–79. <https://doi.org/10.1007/s10988-006-0001-4>
- Richards, Jack C. 1974. Word Lists: Problems and Prospects. *RELJ Journal* 5 (2): 69–84. <https://doi.org/10.1177/003368827400500207>
- Ronald, James. 2002. "L2 Lexical Growth through Extensive Reading and Dictionary Use: A Case Study." In *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002*. Vol 2, edited by Anna Braasch and Claus Povlsen, 765–772. Copenhagen.
- Rundell, Michael. 1988. "Changing the Rules: Why the Monolingual Learner's Dictionary Should Move Away from the Native Speaker Tradition." In *ZürichLEX' 86 Proceedings*, edited by Mary Snell-Hornby, 127–138. Tübingen: Francke Verlag.
- Rundell, Michael. 2006. "Learner's Dictionaries." In *Encyclopedia of Language and Linguistics*, 2nd ed., edited by Keith Brown, Anne Anderson, Laurie Bauer, Margie Berns, Jim Miller, and Graeme Hirst, 739–743. Oxford: Elsevier. <https://doi.org/10.1016/B0-08-044854-2/00427-2>
- Rundell, Michael. 2008a. "Recent Trends in English Pedagogical Lexicography." In *Practical Lexicography A Reader*, edited by Thierry Fontenelle, 221–243. Oxford: Oxford University Press.
- Rundell, Michael. 2008b. "More Than One Way to Skin a Cat: Why Full-Sentence Definitions Have Not Been Universally Adopted." In *Practical Lexicography A Reader*, edited by Thierry Fontenelle, 197–209. Oxford: Oxford University Press.
- Rundell, Michael. 2015. "From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions." *Lexikos* 25: 301–322. <https://doi.org/10.5788/25-1-1301>
- Russo, John P. 1989. *I.A. Richards. His Life and Work*. New York: Routledge.
- Sager, Juan, and Marie-Claude L'Homme. 1994. "A Model for the Definition of Concepts: Rules for Analytical Definitions in Terminological Databases." *Terminology*, Vol. 1 (2): 351–373. <https://doi.org/10.1075/term.1.2.08sag>
- Sager, Juan, and Kyo Kageura. 1994. "Concepts Classes and Conceptual Structures: Their Role and Necessity in Terminology." *ALFA: Terminology and Special Linguistics* 7 (8): 191–216.
- Sager, Juan. C., and Augustin Ndi-Kimbi. 1995. "The Conceptual Structure of Terminological Definitions and their Linguistic Realisations: A Report on Research in Progress." *Terminology* 2 (1): 61–85. <https://doi.org/10.1075/term.2.1.04sag>
- San Martin, A., and Marie-Claude L'Homme. 2014. "Definition Patterns for Predicative Terms in Specialized Lexical Resources." In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. (Reykjavik, Iceland) European Language Resources Association (ELRA), May 2014, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis.
- Santini, Marina, and Serge Sharoff, eds.. 2007. *Proceedings of the CL 2007 Colloquium: Towards a Reference Corpus of Web Genres*. Birmingham, UK.

- Seppälä, Selja. 2015. "An Ontological Framework for Modeling the Contents of Definitions," *Terminology* 21 (1): 23–50. <https://doi.org/10.1075/term.21.1.02sep>
- Seppälä, Selja, Alan Ruttenberg, and Barry Smith. 2016. "The Functions of Definitions in Ontologies." In *Formal Ontology in Information Systems: Proceedings of the 9th International Conference (FOIS 2016), July 6–9 2016*, Frontiers in Artificial Intelligence and Applications, vol. 283, edited by Roberta Ferrario and Werner Kuhn, 37–50. Annecy, France: IOS Press.
- Schmitt, Norbert. 2000. *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.
- Schmitt, Norbert, and Diane Schmitt. 2014. "A reassessment of frequency and vocabulary size in L2 vocabulary teaching." *Language Teaching* 47 (04): 484–503. <https://doi.org/10.1017/S0261444812000018>
- Sutarsyah, Cucu, Nation, I. S. Paul, and Graeme Kennedy. 1994. "How Useful is EAP Vocabulary for ESP? A Corpus Based Study." *RELC Journal* 25 (2): 34–50. <https://doi.org/10.1177/003368829402500203>
- Seretny, Anna. 1998. *Definicje i definiowanie*. Cracow: Zakład Wydawniczy "NOMOS".
- Schul, James E., and Gregory E. Hamot. 2011. "An Engaged Pragmatist: Uncovering and Assessing Ernest Horn's View of Moral Education." *Journal of Social Studies Research* 35 (2): 277–298.
- Shaw, Claire. 2013. *Is English Still the Dominant Language of Higher Education? – Live Chat*. *The Guardian*. Accessed September 2, 2017. <https://www.theguardian.com/higher-education-network/2013/feb/13/english-language-international-higher-education>
- Simpson, John A. 1990. "English Lexicography After Johnson to 1945." In *Wörterbücher Dictionnaires Dictionnaires. An International Encyclopedia of Lexicography*. 2 vol., (1989–91), edited by Franz J. Hausmann, Oskar Reichmann, Herbert E. Wiegand, and Ladislav Zgusta, 1953–1966. Berlin: Walter de Gruyter.
- Sinclair, John M. 1984. "Naturalness in Language." In *Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research*, edited by Jan Aarts and Willem Meijs, 203–10. Amsterdam: Rodopi.
- Sinclair, John M. ed.. 1987. *Looking Up: An account of the COBUILD Project in lexical computing and the development of the Collins COBUILD English Language Dictionary*. London: Collins ELT.
- Sinclair, John M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sin-Wai, Chan. 2015. "Computer-Aided Translation: Major Concepts." In *Routledge Encyclopedia of Translation Technology*, edited by Sin-Wai Chan, 32–67. London: Routledge.
- Smith, Richard C. ed.. 2003. "Introduction." In *Teaching English as a Foreign Language, 1912–1936. Pioneers of ELT. Volume III Michael West*, ix–xxvi. London and New York: Routledge.
- Smith, Richard C. 2015. *Lawrence Faucett's Life and Career*. Accessed March 6, 2018. https://warwick.ac.uk/fac/soc/al/research/collections/elt_archive/halloffame/faucett/life
- Snider, James G., and Charles E. Osgood, ed.. 1969. *Semantic Differential Technique: A Sourcebook*. Chicago: Aldine.
- Sobkowiak, Włodzimierz, and Marek Kuczyński. 2002. "Phonetics and Ideology of Defining Vocabularies." In *EURALEX' 2002 Proceedings, Copenhagen, 13–17 August 2002*, edited by Anna Braasch and C. Povlsen, 495–502. Copenhagen: Center for Sprogteknologi.
- Staehr, Lars S. 2009. "Vocabulary Knowledge and Advanced Listening Comprehension in English as a Foreign Language." *Studies in Second Language Acquisition* 31 (4): 577–607. <https://doi.org/10.1017/S0272263109990039>

- Stein, Gabriele. 1979. "The Best of British and American Lexicography." *Dictionaries* 1:1–23. <https://doi.org/10.1353/dic.1979.0004>
- Stein, Gabriele. 1989. "Recent Developments in EFL Dictionaries." In *Learners' Dictionaries: State of the Art*, edited by Makhan L. Tickoo, 10–41. Singapore: SEAMEO Regional Language Centre.
- Stein, Gabriele. 2002a. *Better Words*. (selected essays by G. Stein on Dictionary Criticism). Exeter: University of Exeter Press.
- Stein, Gabriele. 2002b. *Developing Your English Vocabulary. A Systematic New Approach*. Tübingen: Stauffenburg Verlag.
- Stein, Gabriele. 2016. "Some Thoughts on the Issue of Core Vocabularies. A Response to Vaclav Brezina and Dana Gablasova: 'Is There a Core General Vocabulary?' Introducing the New General Service List." *Applied Linguistics* 38(5): 759–763.
- Sterkenburg, Piet van, ed.. 2003. *A Practical Guide to Lexicography*. Amsterdam: John Benjamins. <https://doi.org/10.1075/tlrp.6>
- Stubbs, Michael. 1986. "Language development, lexical competence and nuclear vocabulary." *Educational Linguistics*, edited by Michael Stubbs, 98–115. Oxford: Blackwell.
- Summers, Della. 1988. "English Language Teaching Dictionaries: Past, Present, and Future." *English Today* 4 (2): 10–16. <https://doi.org/10.1017/S0266078400013274>
- Svensén, Bo. 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- Swanepoel, Piet. 1994. "Problems, Theories and Methodologies in Current Lexicographic Semantic Research." In *EURALEX 1994 Proceedings. Papers submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam (Vrije Universiteit) 1994*, edited by Willy Martin, Willem Meijs, and Margreet Moerland, 11–26. Amsterdam.
- Sweet, Henry. 1899. "The Practical Study of Languages: A Guide for Teachers and Learners." London: J. M. Dent. [Reprinted 2003. *Lexicography Critical Concepts*. vol. 2, edited by Reinhard R.K. Hartmann, 257–268. London: Routledge.]
- Ščerba, Lev. 1995. "Towards a General Theory of Lexicography." (translated by Farina, D.) *International Journal of Lexicography* 8 (4): 305–349.
- Taylor, John R. 1995. *Linguistic categorization. Prototypes in linguistic theory*. Oxford: Clarendon Press.
- Thorndike, Edward L. 1921. *The Teacher's Word Book*. Columbia University: Teachers College.
- Thorndike, Edward L., and Irving Lorge. 1944. *The Teacher's Word Book of 30,000 Words*. Oxford: Bureau of Publications, Teachers Co.
- Thouin, Benoît. 1982. "The Meteo System." In *Practical Experience of Machine Translation: Proceedings of Translating and the Computer 1981*, edited by Veronica Lawson, 39–44.
- Tickoo, Makhan L. 2008. *Harold E. Palmer: From Learner-teacher to Legend*. Hyderabad: Orient Longman.
- Tomaszczyk, Jerzy. 1979. Dictionaries: Users and Uses, *Glottodidactica* 12: 103–19.
- Tono, Yukio. 1989. "Can a Dictionary Help One Read Better?" In *Lexicographers and their Works*. (Exeter Linguistic Studies Volume 14), edited by Gregory James, 192–200. University of Exeter.
- Trask, Robert L. 1996. *Historical Linguistics*. London: Arnold.
- Trier, Jost. 1931. *Der Deutsche Wortschatz im Sinnbezirk des Verstandes: Die Geschichte eines Sprachlichen Feldes I. Von den Anfängen bis zum Beginn des 13. Jhdts*. Heidelberg: Winter.

- Ungerer, Friedrich, and Hans-Jörg Schmid. 2006. *An Introduction to Cognitive Linguistics*. Harlow: Pearson Longman.
- VOA Special English Word Book A–Z. 2009. Washington, D.C. www.VOASpecialEnglish.com.
- WaCKy. 2017. *The ukWaC unigrams (lemmas)*. Accessed May 10, 2017. https://wacky.sslmit.unibo.it/doku.php?id=frequency_lists.
- Waring, Rob, and Misako Takaki. 2003. "At What Rate Do Learners Learn and Retain New Vocabulary from Reading a Graded Reader?" *Reading in a Foreign Language* 15 (2): 130–163.
- Webb, Stuart. 2007. "The Effects of Repetition on Vocabulary Knowledge." *Applied Linguistics* 28 (1): 46–65. <https://doi.org/10.1093/applin/aml048>
- Webb, Stuart, Paul I. S. Nation. 2008. "Evaluating the Vocabulary Load of Written Text." *TESOLANZ Journal* 16: 1–10.
- Weinreich, Uriel. 1980. *On Semantics*. Edited by William Labov and Beatrice S. Weinreich. Philadelphia: University of Pennsylvania Press. <https://doi.org/10.9783/9781512819267>
- Wekker, Herman, and Phil Hyams. 1979. "De Longman Dictionary of Contemporary English." Review Article. *Levende Talen* 342: 406–417.
- West, Michael. 1926. *Bilingualism (with Special Reference to Bengal)*. Calcutta: Government of India. (Reprinted in: Richard C. Smith, ed. 2003. Vol. 3, 1–392).
- West, Michael. 1927a. *The Construction of Reading Material for Teaching a Foreign Language*. London: Oxford University Press.
- West, Michael. 1927b. *The New Method Readers for Students of English*. Longmans, Green and Co.
- West, Michael. 1935. *Definition Vocabulary*. Bulletin no. 4 of the Department of Educational Research. Toronto: University of Toronto Press.
- West, Michael. 1953. *A General Service List of English Words with Semantic Frequencies and a Supplementary Word-List for the Writing of Popular Science and Technology (GSL)*. London: Longmans.
- West, Michael. 1954. "Vocabulary Selection and the Minimum Adequate Vocabulary." *English Language Teaching* 8 (4): 121–126. <https://doi.org/10.1093/elt/VIII.4.121>
- West, Michael. 1960. *Teaching English in Difficult Circumstances: Teaching English as a Foreign Language with Notes on the Technique of Textbook Constructions*. London: Longmans, Green.
- West, Michael, E. Swenson, K. M. Fawkes, F. L. Russell, and Wilf J. de Magellanes. 1934. *A Critical Examination of Basic English*. Bulletin of Department of Educational Research, University of Toronto, 2: 1–53.
- Whitcut, Janet. 1978. "Lexicography in Controlled Vocabulary." Paper presented at the AILA conference. Montreal 1978.
- Whitcut, Janet. 1986. "The Training of Dictionary Users." In *Lexicography An Emerging International Profession*, edited by Robert Ilson, 111–122. London: Manchester University Press.
- Whitcut, Janet. 1988. "Lexicography in Simple Language." *International Journal of Lexicography* 1 (1): 49–55. <https://doi.org/10.1093/ijl/1.1.49>
- Wiegand, Herbert E. 1994. "Elements of a Theory Toward a So-called Lexicographic Definition." In Wiegand, Herbert E. 1999. *Semantics and Lexicography. Selected Studies (1976–1996)*, 203–282, edited by Antje Immken and Werner Wolski. Tübingen: Max Niemeyer.
- Wingate, Ursula. 2002. *The Effectiveness of Different Learner Dictionaries. An Investigation into the Use of Dictionaries for Reading Comprehension by Intermediate Learners of German*. Tübingen: Max Niemeyer Verlag.

- Widdowson, H. G. 1968. "The Teaching of English Through Science." In *Language in Education*, edited by Julian Dakin, Brian Tiffen, and H. G. Widdowson. London: Oxford University Press.
- Wierzbicka, Anna. 1972. *Semantic Primitives*. Frankfurt.
- Wierzbicka, Anna. 1985. *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma Publishers, Inc.
- Wierzbicka, Anna. 1996. *Semantic Primes and Universals*. Oxford: Oxford University Press.
- Wierzbicka, Anna. 2003. *Cross-Cultural Pragmatics: The Semantics of Human Interaction*. 2nd ed. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110220964>
- Wierzbicka, Anna. 2014. *Imprisoned in English. The Hazards of English as a Default Language*. New York: Oxford University Press.
- Wilks, Yorick, Dan Fass, Cheng-ming Guo, James E. McDonald, Tony Plate, and Brian M. Slator. 1988. "Machine Tractable Dictionaries as Tools and Resources for Natural Language Processing." In *Proceedings of the 12th International Conference on Computational Linguistics (Coling-88) Budapest*, Vol. 2, 750–755, Association for Computational Linguistics. <https://doi.org/10.3115/991719.991789>
- Williams, Christopher. 2004. "Legal English and Plain Language: An introduction." *ESP Across Cultures* 1: 111–124.
- Wojcik, Richard H. 2006. "Controlled Languages." In *Encyclopedia of Language and Linguistics*, 2nd ed., edited by Keith Brown, Anne Anderson, Laurie Bauer, Margie Berns, Jim Miller, and Graeme Hirst, 136–142. Oxford: Elsevier. <https://doi.org/10.1016/B0-08-044854-2/05081-1>
- Xiao, Richard. 2008. "Well-known and influential corpora." In *Corpus Linguistics: An International Handbook*, Vol. 1, edited by Anke Lüdeling and Merja Kytö, 383–456. Berlin: Walter de Gruyter.
- Xu, Hai. 2012. "A Critique of the Controlled Defining Vocabulary in Longman Dictionary of Contemporary English." *Lexikos* 22: 367–381. <https://doi.org/10.5788/22-1-1013>
- Xue, Guoyi, and Paul I. S. Nation. 1984. "A University Word List." *Language Learning and Communication* 3 (2): 215–229.
- Youngblood, Alison M., and Keith S. Folse. 2017. "Survey of Corpus-Based Vocabulary Lists for TESOL Classes." *MEXTESOL Journal* 41 (1): 1–15.
- Zakaluk, Beverley L., and S. Jay Samuels. 1996. "Issues Related to Text Comprehensibility: The Future of Readability." *Revue Québécoise de Linguistique* 25 (1): 41–59. <https://doi.org/10.7202/603126ar>
- Zarechnak, Michael. 2000. "The early days of GAT-SLC." In *Early Years in Machine Translation. Memoirs and Biographies of Pioneers*, edited by W. John Hutchins, 111–128. Amsterdam: John Benjamins. <https://doi.org/10.1075/sihols.97.11zar>
- Zipf, George K. 1935. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Boston: Houghton-Mifflin.
- Żmigrodzki, Piotr. 2003. *Wprowadzenie do leksykografii polskiej*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Index

A

abstract words 173
 Academic Word List 55
 allowable
 meanings 189, 205, 210
 words 179, 181, 190, 191,
 198, 202, 203, 204,
 208, 248
 ambiguity 19, 65, 68, 76, 115,
 116
 American Brown Corpus 50
*American Heritage Word
 Frequency Book* 48
 analytical definition 91, 92,
 179, 211

B

Bank of English 50, 51
 Basic English 29, 31, 32, 49, 50,
 62, 63, 69, 85
 British National Corpus 43,
 50, 52, 58, 60, 61, 72, 73, 124,
 131, 142, 145, 187

C

Cambridge English
 Lexicon 49
 Cambridge International
 Corpus 52
 Carnegie Conference 33
 COBUILD corpus 51, 124
 collocability 72
 collocational range 72, 73
 Common Core Vocabulary 56,
 128, 164, 165
 comprehensibility
 (of definitions) 118, 119, 169,
 201, 216
 computer (in dictionary
 compilation) 112, 198, 213
 concrete words 17, 173
 controlled language 29, 62,
 63, 65
 core defining vocabulary 128,
 131, 163, 249
 core meaning 56, 75, 78, 101

core vocabulary 56, 69, 70, 71,
 74, 82
 coreness 69, 70, 73, 75
 credit-number 14
 culture-specific words
 163, 175

D

definiendum 90, 91, 92, 140
definiens 92, 140
 what is a definition? 90, 92
 digital dictionaries 213, 214
 direct method 104

E

Eldridge, R.C. 13, 19, 69

F

Faucett, Lawrence 15, 16, 17, 33,
 34, 39, 59, 164
 Faucett-Maki list 34
 formal productivity 71
 full-sentence definition 4, 91,
 140, 196

G

General Service List 22, 33, 34,
 35, 36, 38, 39, 40, 41, 42, 43,
 44, 45, 47, 50, 51, 53, 54, 55,
 56, 58, 59, 60, 61, 62, 69, 70,
 112, 114, 164, 166, 249
genus proximum 90, 174
 grammar-translation
 method 104
Grammar of English Words 22

H

hard words 93, 95, 122
 Horn, Ernest 15, 16, 18, 28,
 34, 48
 Hornby, A.S. 2, 21, 23, 34, 89,
 100

I

Interim Report 16, 21, 22, 34,
 38, 39

international words 31, 166
 IRET 19, 20, 100, 107

L

lemma 55, 56, 57, 58, 175
 lexeme 55, 59, 87, 158, 167
 lexical set 71, 72
 lexical simplicity (of
 definitions) 137
 lexical unit 20, 36, 57, 91
 LOB corpus 50, 51, 53, 55, 59,
 60, 61

M

machine translation 62, 63, 65,
 66, 67, 69, 87
 Maki, Itsu 15, 16, 34, 59, 164
 Minimal English 78, 85
 minimum adequate
 vocabulary 24
 Murray, James 70

N

Natural Semantic
 Metalanguage 47, 78, 82
 New Academic Vocabulary
 List 57
 New General Service List 58
New Method Readers 23, 25

O

objective approach 13
 Ogden, Charles K. 21, 28, 29,
 30, 31, 34, 63
 opaque
 combinations of words 177
 derivatives 179, 203
 phrasal verbs 167, 180, 217
 words 191, 204

P

Palmer, Harold 17, 19, 20, 21,
 22, 23, 28, 31, 32, 33, 34, 59,
 74, 107
 plain English 93, 102, 104
 Plain English Campaign 103

precision (of definitions) 83,
102, 185, 199, 221
primitive-based dictionary 82
prototype theory 75, 77

R

range 7, 14, 27, 35, 42, 43, 53, 57,
124, 143
receptive vocabulary load 7,
132, 134, 135, 145, 148, 149, 153,
168, 169, 171, 216, 220, 227,
228, 233, 234, 236, 237, 241,
245, 247, 248, 250
register 14, 61, 74
restricted collocation 116

S

Semantic Count of English
Words 38
semantic productivity 71
sense (in a dictionary
entry) 36, 38, 50, 73, 75, 76,
77, 89

simple ideas 96
Six Thousand Common English
Words 13, 69
stenography 12, 13
subjective approach 18
superordinateness 71
synonym definition 90, 91,
92, 134, 174, 221, 229, 242,
244, 245

T

Teacher's Word Book 14, 15, 21
telegraphese 100
text coverage 53, 60, 61, 144,
145, 148
Thorndike, Edward L. 7, 14, 15,
16, 17, 18, 19, 21, 24, 26, 28,
34, 38, 48, 50, 54, 69
Thousand-Word English 21, 22
Threshold Level English 48, 59

U

University Word List 53, 54

W

West, Michael 1, 2, 3, 15, 19,
21, 23, 24, 25, 26, 27, 28, 31,
32, 33, 34, 37, 38, 39, 40, 42,
47, 58, 59, 72, 74, 89, 104, 105,
106, 107, 112, 113, 114, 117, 134,
164, 169, 215, 217, 247, 248
word class (in a defining
vocabulary) 41, 173
word family 21, 36, 37, 53, 55,
56, 57, 58, 59, 60, 61, 124,
131, 142, 143, 144, 146, 148,
149, 152, 153, 164, 168, 169,
186, 204, 205, 220, 241, 247,
248, 250
word sense
disambiguation 76, 77
word form 17, 55, 58, 59, 143,
158
WordNet 77

Z

Zipf, George K. 16

This book investigates an important but under-researched aspect of dictionary making: the use of a controlled vocabulary in definitions. The main concern of the author is the role of a definition vocabulary in how foreign learners understand and perceive dictionary definitions. The author takes the reader through a detailed historical account of controlled vocabularies and examines definitions in a range of English dictionaries with respect to their vocabulary loads. He performs a series of experiments with university students to reveal merits and shortcomings of restricted vocabularies. This monograph has been written with the aim to fill a gap in the literature on defining vocabulary. It is intended for lexicographers, dictionary editors, course designers, teachers, and students, as well as anyone who wishes to explain words in an intelligible way.

ISBN 978 90 272 0859 0



John Benjamins Publishing Company