

Phonetics and Ideology of Defining Vocabularies¹

Włodzimierz Sobkowiak

School of English, Adam Mickiewicz University
al. Niepodległości 4, 61-874 Poznań, Poland

Marek Kuczyński

Teachers' Training College of English
University of Zielona Góra
ul. Wojska Polskiego 69, 65-762 Zielona Góra, Poland

Abstract

The controlled defining vocabularies of LDOCE3 and CIDE have been tested for phonetic difficulty and ideological bias. The research hypotheses were that: (1) they are on the whole not phonetically easier than the 'ordinary' lexicon of English, and (2) they are ideologically biased in that they tend to reflect the realm of material concepts better than spiritual. Both hypotheses failed: the two vocabularies are in fact phonetically easier than the reference lexicon (significant at $p < .01$), and the slight bias towards the 'appearance&body' concepts in both vocabularies is statistically nonsignificant. Didactic ramifications in the EFL context as well as the lexicographic import of the findings are discussed.

Introduction

Considering the current status of the controlled defining vocabulary in EFL lexicography — that of an essential feature of a modern learner's dictionary — it is surprising to find rather little scholarly treatment of it, by metalexicographers, EFL methodologists or linguists. And yet, if it is indeed the case that (a) EFL learners actually learn anything from dictionaries they use [rather than only look them up for their immediate lexical needs and quickly forget what they found; see Summers 1988 for discussion], and that (b) in their majority they want to find out about the meaning of an unknown word (as is amply attested in the available dictionary-use literature), the pedagogical role of definitions in EFL monolingual dictionaries cannot be overestimated. As far as their lexical component is concerned, then, the choice of the deployed vocabulary is quite crucial for the learner's vocabulary acquisition as a whole.

In this paper we have chosen to address two issues of rather direct relevance to the structure and function of defining vocabularies, as used in monolingual dictionaries of English for foreign learners (EFL). The two issues are located at the two extremes of linguistic investigation: phonetics on one and ideology on the other. They can be phrased as follows:

- **Is defining vocabulary phonetically easier than 'ordinary' English vocabulary?**
- **Is defining vocabulary balanced in its representation of material vs spiritual concepts?**

In the following we will take a closer look at whether the two issues have been addressed by the designers of defining vocabularies in two of the leading four British EFL monolingual dictionaries: *Longman Dictionary of Current English* [LDOCE; Summers 1995] and

Cambridge International Dictionary of English [CIDE; Procter 1995]. We will first briefly look at what the editors explicitly declare about their defining vocabularies (in section 1), then we will dissect their phonetic structure (section 2; Sobkowiak) and ideological assumptions (section 3; Kuczyński). Finally conclusions and postulates will be presented for EFL pedagogy and lexicography.

Defining vocabularies: what we learn about them

In Appendix B15 to the second, 1987, edition of LDOCE [14th impression, 1992] we read: "The Defining Vocabulary has been carefully chosen after a thorough study of all the well-known frequency lists of English words. Furthermore, only the most common and "central" meanings of the words in the list have actually been used in definitions". Later, it is emphasised that parts of speech have also been controlled in case of homographs, that only fully transparent compounds have been used (e.g. *businessman*), that affixes from the appended list can be combined with stems to form derivatives, and that only the synonymless phrasal verbs have been allowed (e.g. *give up, take off*). We can learn from other sources that the LDOCE3 defining vocabulary was also matched against the Longman Learners' Corpus containing about ten million words written by EFL learners from all over the world. "Words that are not frequent on the Longman Learners' Corpus are not included in the Longman Defining Vocabulary" [Fox 1998:12].

In turn, *CIDE* says, introducing the full list "of under 2000 words"ⁱⁱ on page 1702: "The words in this list have been carefully chosen, according to these principles", which are then nicely laid-out in a flower-like graph, and include: semantic transparency, commonness, definitional usefulness, avoidance of ambiguity, and others. Notice that the obvious superordinate principle in both LDOCE and CIDE is making it easier to (non-advanced) learners to understand the definitions in the dictionary, at least on the lexical level. Thus, controlled defining vocabulary is really simplified defining vocabulary.

In neither LDOCE or CIDE is there any mention of compilers' control over the phonetics or ideology of their defining vocabularies. The original hunch of the present authors was that there was indeed no effective control over the two aspects of defining vocabularies, with all the ramifications of this neglect for the use of the dictionaries in an EFL pedagogical context. It was felt that, on the one hand, defining vocabularies are not phonetically easier than the (properly operationalised) 'average' lexicon, and on the other — that defining vocabularies might be ideologically imbalanced in that they would tend to reflect the realm of material concepts better than spiritual.

Phonetics of defining vocabularies

Is defining vocabulary supposed to be easy in terms of pronunciation? Prima facie, the answer would appear to be a resounding yes: if lexicographers are aiming at global 'user-friendliness' of their defining vocabulary, they should certainly also make it phonetically friendly; the more so as more and more evidence exists that pronounceability, especially in the case of beginning foreign language learners, is an important factor in lexical access and selection as well as word perception, comprehension and retention [see Nation 1990:36; Eckman 1981; Laufer 1990 and 1997; and Singleton 1999:138ff for an overview and references].

While I am not aware of any directly relevant research [phonetics is conspicuously missing from Michiels & Noël 1984, Herbst 1986, Neubauer 1987 or Jansen, Mergeai & Vanandoroye's 1987], it stands to reason that phonetically difficult words in dictionary definitions will tend to impede the reading and understanding process, particularly in those learners who continue to vocalise or articulate subvocally in silent reading. And, according to Gibson and Levin [1980:342]: "it is perfectly certain that the inner hearing or pronunciation, or both, of what is read, is a constituent part of reading by far the most of people, as they ordinarily and actually read".

The hypothesis

Thus, one would expect that defining vocabularies are phonetically easier/simpler than the lexical average (properly operationalised), even if no such simplification is hinted at in the front matter of the respective dictionaries. However, in my long learner's and then researcher's experience I have been aware of a number of lexical items used in dictionary definitions which seem to me anything but phonetically easy. Take a short selection from CIDE: advantage, although, behaviour, discourage, journey, theatre, uncontrollable. Further prompted by the awareness of the thoroughly marginal status of phonetics in current (meta)lexicographic research, which is amply documented in the first chapter of my book [Sobkowiak 1999], I formulated a working hypothesis that defining vocabularies are not phonetically easier/simpler than the lexical average. To test this hypothesis, the following operationalising steps have been made:

- The computer-readable LDOCE3 defining vocabulary list has been retrieved from <http://www.sys.uea.ac.uk/~jrk/conlang.dir/LongmanVocab.html>. This lemmatised list counts exactly 2197 entries. The defining vocabulary of CIDE has been scanned and OCRed from the appendix to the hard-copy version of *CIDE* [Procter 1995; pages 1702-1707]. After some editing (dropping multiword entries and single-letter entries like *E* for 'east') the list came to 3550 records.
- "Phonetically easier" has been operationally defined as "having a statistically significant lower mean score on the L1-sensitive phonetic difficulty index", as explained in Sobkowiak 1999.
- "Lexical average" has been operationally defined as "a lemmatised frequency-matched subset of the word-list derived from the computer-readable version of OALDCE", as described in Mitton 1986 and 1992. Notice that this does not mean the defining vocabulary list from OALDCE.

Frequency matching was necessary because phonetic difficulty is indirectly related to word frequency. More frequent words tend to be shorter (Zipf) and thus phonetically easier to learners on two counts: first, shorter phonetic strings are easier to pronounce than longer ones, *ceteris paribus*; second, the shorter the word is, the fewer phonetically or grapho-phonetically difficult points it will contain. It would thus be misleading to compare a defining vocabulary against a frequency-unmatched sample of the lexicon.

The procedure

The original 2197 LDOCE3 and 3550 CIDE defining vocabulary items were matched against the lemmatised frequency list of 6318 words (those with frequencies of 800 upwards) derived by Adam Kilgarriff from the 100-million-word-long BNC corpus [ftp://ftp.itri.bton.ac.uk/pub/bnc/; see also Kilgarriff 1997]. As it turned out, 182 defining words of LDOCE and 1200 words of CIDE do not occur on the BNC-LEM list, i.e. they are outside the group of the most frequent 6318 words in English.

Of the 2350 CIDE words listed among BNC lemmas I took the most common 2015, as ranked by Kilgarriff, to match with the 2015 LDOCE defining words also listed there (2197-182=2015). The two word-lists showed a remarkably similar average BNC-LEM frequency of 28,359 for LDOCE and 28,561 for CIDE, i.e. about 285 occurrences per one million running words. To obtain a frequency-matched word-list from OALDCE, the top 2015 BNC-LEM-ranked words were taken from it. After all this normalisation, the frequency distribution of the difficulty index was checked for normalcy over all three lists (χ^2 test was used). As the lists turned out not to be normally distributed in terms of phonetic difficulty (indeed they are all rather heavily skewed towards the easy end of the spectrum) the Mann-Whitney rank test was used to see whether and to what extent the two defining vocabulary lists diverge from reference in terms of phonetic difficulty. The value of z turned out to be -2.76 for LDOCE3 and 0.9 for CIDE. This means that LDOCE3 defining vocabulary is indeed phonetically easier (has a generally lower phonetic difficulty index score; hence the negative z) than chance, as operationalised above. This conclusion is probable at the level of 99% or better.

As far as CIDE is concerned, the result is of course statistically nonsignificant, and the positive value suggests that CIDE's defining vocabulary might indeed be a bit more phonetically difficult than chance, which would be rather paradoxical. After stripping derivatives from CIDE, however, I was left with 1715 words, which were again ranked and subjected to the Mann-Whitney test. The result this time came out at $z=-3.6$, which shows that the non-derived CIDE defining vocabulary is indeed, with very high probability, phonetically easier than chance.

The results

The overall conclusion is that defining vocabularies are, after all, significantly phonetically easier than the frequency-matched portions of the reference lexicon, here treated as chance level. The working hypothesis claiming no such effect is thus refuted. This does not necessarily show that editors of the two learners' dictionaries of English have exercised effective control over this important aspect of the defining vocabulary's structure. Let alone that they have intentionally phonetically simplified their lists in any way (e.g. by phonetically sensitive selection).

Be it as it may, the conclusion is rather heartening. Phonetically difficult defining vocabularies would bring about a number of unwelcome consequences: they might impede vocabulary acquisition, foster phonetic error fossilization, create cognitive discomfort in

learners, who are apt to assume that if vocabulary is simplified for their sake, it would be simplified overall.

Ideology of defining vocabularies

This part of the paper aims at examining the degree to which the two defining vocabulary lists (LDOCE&CIDE) equip the learner with the lexical tools necessary for exchanging knowledge concerning human appearance, health, personality and spirituality. These being central to the expression of human self-awareness and the ability to perceive others, a relative balance between the corresponding sub-lexicons should be expected. This is a question the answer to which is by no means obvious, for the author has recently found out that the absence of such a balance is in fact the case in English coursebooks [Kuczyński 1999].

Foreign language learning and cognition – a summary

The decision on the lexical content of a defining vocabulary list is more crucial than may at first seem [see Nation 1990, Carter and McCarthy 1988]. First, selection determines the areas of human experience which may become the subject of verbal communication with another person speaking the target language. Including the word life in the list will probably result in exposing the learner to this item in numerous definitions, which almost certainly ensures acquisition. This, in turn, will enable the learner to exchange information or views related to this concept. Second, if language is believed to be related to thinking and cognition in general, selection would consequently be thought to at least have a bearing upon the learner's awareness of that fraction of reality which is contained in a concept and which has a linguistic label attached to it (but it is fair to admit in this connection that there is little in the literature on the relationship that exists between one's second/foreign language and one's cognition or thinking abilities as such).

Thus, a defining vocabulary list designer favours some concepts which will be activated and disfavours those that will not. S/he could affect the directions in which the learner's cognitive space will expand. It needs to be added here that all the above is only true to the extent to which students' learning time is taken up by looking up meanings in monolingual dictionaries. There are, of course, other sources of the vocabulary which is acquired in various learning and real-life situations. The issues considered here are not regarded as determining the student's cognitive structure but only contributing to it. It is the quality of this contribution that is subjected to examination.

The research question

To what degree do LDOCE and CIDE contribute to an equal development of those areas in the learners' cognitive system which are responsible for storing information, processing as well as expressing judgements concerning the two realms of human experience: appearance&body as well as personality&spirituality?

The selection of these areas was meant to emphasise the contrast between the material and non-material facets of our existence. The assumption (which had developed as a result of obtaining data from coursebooks) made by the author before embarking on this experiment was that the areas would not receive equal treatment by the list designers. Kuczyński [1999]

found that the authors of some coursebooks for English overexposed learners to content related to the body and appearance, and, at the same time, neglected the areas of human personality and spirituality.

In order to answer the research question, the following steps will be taken:

- Two subsets (appearance&body, personality&spirituality) of LDOCE and CIDE will be identified in an attempt to organise the relevant LDOCE and CIDE lexical items into two corresponding sub-lexicons, on the basis of the author's perception and intuition (with all the admitted drawbacks of this method), as shown in the table below:

appearance&body	personality&spirituality
states of the body	states of the mind/spirit
qualities of the body	characteristics of personality
parts of the body	personality/spirituality components
qualities/components of appearance	attitudes

Table 1: Vocabulary subset selection criteria

- The sizes of the two subsets will be measured by counting the number of items which constitute them. This alone offers an insight into whether there is a relative balance between the extent to which the two areas are represented lexically in the lists. A better solution would certainly be to analyse the definitions, but such a task would need decades to complete. For the sake of feasibility, only the lists were processed for numerical data. A similar procedure was adopted in Sobkowiak's section of this paper for studying the correspondence in terms of phonetic difficulty.
- The significance of the data obtained will be validated by means of the χ^2 test.

Findings

	body & appearance	personality & spirituality	χ^2
LDOCE	122	100	2.18 (n.s.)
CIDE	235	217	0.72 (n.s.)

Table 2: The two sub-lexicons in LDOCE and CIDE

Taking into account the arbitrariness of the individual decisions in compiling the sub-lexicons (which implies the need for a certain margin of error within which the arbitrariness-related contamination could be accounted for), it may be said that the findings do not seem to reflect a dramatic misrepresentation in either the first or the second list. That the *appearance&body* sub-lexicon compiled in the particular way outnumbers the other in both lists is a matter of fact; how much this mismatch is a reflection of an ideal, non-contaminated-by-arbitrariness, size of the sub-lexicons would not be easy to determine within the framework adopted for the purpose of this experiment. The research, the way it

was constructed, would entitle the author to draw conclusions rejecting the claim that the sub-lexicons are represented equally if the figures differed by more than they in fact do.

This is an interesting observation, because it stands in contrast with what has been observed by Kuczyński [1999] in relation to such representativeness in English coursebooks. It was found that, roughly speaking, the personality-related lexicon was significantly underrepresented if compared with the body-oriented lexicon (by nearly a half). The present study places that observation within a more interesting, broader perspective: the existing defining vocabularies in terms of both the body and the mind have not been shown to vary as dramatically as they do in coursebooks. This means that coursebooks do not tend to reflect the defining language which has to do with describing the two facets of human existence. LDOCE and CIDE may be feared to slightly disregard the personality/spirituality area when compared to the appearance/body sub-lexicon, but the findings which may suggest this are not significant statistically.

General conclusions

We have shown in this paper that, despite our working hypotheses, defining vocabularies of two of the "big four" learners' dictionaries of English as a foreign language are phonetically easier than a frequency-matched sample of the English lexicon, and that they project an ideologically (roughly) balanced view of reality, at least along the dimension chosen for scrutiny in section 3. What we have not demonstrated, as this was not the aim of the present study, is whether these properties of defining vocabularies originated by deliberate design and conscious effort on the part of dictionary makers, or rather are due to their skill, experience and intuition as lexicographers, educators and native speakers of English. We tend to entertain a lingering feeling that the latter is indeed the case, and we could not find arguments either way in the available literature of the field. Our feeling is informed by a number of factors, not the least being explicit comments and remarks coming from dictionary compilers showing that they effectively control only the middle spectrum of linguistic structure in defining vocabularies, i.e. that located between morphology and pragmatics.

If our observation is indeed true, that is if lexicographers achieve a pedagogically desirable state of affairs as far as some parameters of defining vocabularies are concerned without dedicated research and solely on the basis of their intuition, they should at the same time be commended for their art and reprimanded for their (lack of) science. Luckily, the latter can easily be corrected. It is high time, we believe, that it indeed be corrected. Defining vocabularies, as was rather forcefully argued by Kuczyński above, are not just a tool of the trade (for dictionary makers) or a welcome aid in lexical look-up (for dictionary users). They strongly affect the whole process of foreign vocabulary learning and teaching, the process which is currently believed to be at the very heart of mastering a good communicative command of a foreign language. And yet, compared to the whole volume of the existing literature on dictionary design and use, the subset devoted to defining vocabularies appears minute, out of all proportion to their past and future role in lexicography. With this paper we hope to have changed this proportion a bit in the right direction.

Endnotes

- ⁱ Thanks go to Tadeusz Piotrowski for his comments on the preliminary version of the paper. The original version, which counted over 40 pages, had to be dramatically shortened here to meet the requirements of Euralex stylesheet.
- ⁱⁱ The full list is really much longer, if all derivatives are counted; see below. Similarly, the declared "approximately 2000 words" of LDOCE3 defining vocabulary resolves to quite a few more. According to Herbst (1986:105), "the number of words used in LDOCE definitions could be estimated to lie between 5,000 and 10,000".

References

- [Carter and McCarthy 1988] Carter,R. & M.McCarthy (eds). 1988. *Vocabulary and language teaching*. London: Longman.
- [Eckman 1981] Eckman,F.R. 1981. "On predicting phonological difficulty in second language acquisition". *Studies in Second Language Acquisition* 4.18-30.
- [Fox 1998] Fox,C. 1998. "Defining vocabularies". *Longman Language Review* 4.11-14.
- [Gibson and Levin 1980] Gibson,E.J. and H.Levin. 1980. *The psychology of reading*. Cambridge, Mass.: The MIT Press.
- [Herbst 1986] Herbst,T. 1986. "Defining with a controlled vocabulary in foreign learners' dictionaries". *Lexicographica* 2.101-19.
- [Jansen, Mergeai & Vanandoroye's 1987] Jansen,J., J.P.Mergeai & J.Vanandoroye 1987. "Controlling LDOCE's controlled vocabulary". In A.P.Cowie (ed.). 1987. 78-94.
- [Kilgarriff 1997] Kilgarriff,A. 1997. "Putting frequencies in the dictionary". *International Journal of Lexicography* 10.2.135-55.
- [Kuczyński 1999] Kuczyński,M. 1999. "Wartości duchowe w nauczaniu języków obcych" [Spiritual values in FLT]. *Studia Paradyskie IX, Paradyż* 1999.
- [Laufer 1990] Laufer,B. 1990. "Ease and difficulty in vocabulary learning: some teaching implications". *Foreign Language Annals* 23.2. 147-55.
- [Laufer 1997] Laufer,B. 1997. "What's in a word that makes it hard or easy: some intralexical factors that affect the learning of words". In N.Schmitt & M.McCarthy (eds). 1997. 140-55.
- [Michiels& Noël] Michiels,A. & J.Noël. 1984. "The pro's and con's of a controlled defining vocabulary in a learner's dictionary". In R.R.K.Hartmann (ed.). 1984. 385-94.
- [Mitton 1986] Mitton,R. 1986. "A partial dictionary of English in computer usable form". *Literary and Linguistic Computing* 1.214-15.
- [Mitton 1992] Mitton,R. 1992. "A description of a computer-usable dictionary file based on the Oxford Advanced Learner's Dictionary of Current English". [Attached with the software]
- [Nation 1990] Nation,P. 1990. *Teaching and learning vocabulary*. New York: Newbury House.
- [Neubauer 1987] Neubauer,F. 1987. "How to define a defining vocabulary". In R.F.Ilson (ed.). 1987. 49-59.
- [Procter 1995] Procter,P. (ed.). 1995. *Cambridge international dictionary of English*. Cambridge: Cambridge University Press. [CIDE]
- [Singleton 1999] Singleton,D. 1999. *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- [Sobkowiak 1999] Sobkowiak,W. 1999. *Pronunciation in EFL machine-readable dictionaries*. Poznań: Motivex.
- [Summers 1988] Summers,D. 1988. "The role of dictionaries in language learning". In R.Carter & M.McCarthy (eds).1988.111-125.
- [Summers 1995] Summers,D. (ed.). 1995. *Longman dictionary of contemporary English* [3rd ed.]. Harlow: Longman. [LDOCE3]