



University of Brighton

ITRI-97-05 **The hard parts of lexicography**

Adam Kilgarriff

July, 1997

Also to appear in the International Journal of Lexicography

This work was supported by the EPSRC under grant K18931

Information Technology Research Institute Technical Report Series

ITRI, Univ. of Brighton, Lewes Road, Brighton BN2 4GJ, UK
TEL: +44 1273 642900 EMAIL: firstname.lastname@itri.brighton.ac.uk
FAX: +44 1273 642908 NET: <http://www.itri.brighton.ac.uk>

In this squib I report on a small survey in which I asked lexicographers to rank various aspects of their work according to how hard they were.

The lexicographers were the team working on the Third Edition of the Longman Dictionary of Contemporary English (LDOCE3). Each was given a sheet as reproduced in Figure 1.

The breakdown of the tasks into subtasks was based on working practices at Longman and the lexicography literature. The dictionary was corpus-based, and lexicographers examined corpus data, on line, for each word before writing the dictionary entry, hence task 5. The “lexunit” and “collocation” of task 6 are two categories of fixed or semi-fixed multi-word expression defined in the LDOCE3 Style Guide which receive particular typographic treatments in the printed product. The major reason for asking “how often” was so that respondents were discouraged from thinking of rarer tasks as less hard because they spent less mental effort thinking about them in total, even though each difficult case of a rare task may take quite as long as each difficult case of a common task. It is not clear whether the extra question succeeded in this goal.

I received 11 responses, and summed the scores for each task. The results, ordered from “hardest” to “least hard” are presented in Table 1.

Order	Task	Score
1	Finding right wording	28
2	Splitting : identifying the senses	46
3	Priority ... meaning and use	47
4	Multi word items	52
5	Inclusion, senses	76
6	Grammar	77
7	Consistency with definitions elsewhere ...	82
8	Examples	86
9	Inclusion, headwords	87
10	Extent	90
11	Knowing when you have spent long enough ...	94
12	Inflections, variant forms	100
13	Runons	130

Table 1: Tasks ordered by difficulty

The tasks can be divided into analysis and synthesis components. Wording, inclusion and consistency are purely synthetic tasks. They relate to what goes into the dictionary, and not at all to determining how words behave or what they mean. Splitting is largely an analysis task. The lexicographer works out how the word’s behaviour is to be analysed into distinct senses in principle before starting to write, though the number of distinctions to be made will vary with dictionary size and style. Analysis is unaffected by the particularities of any dictionary but synthesis is directly and obviously related to the particular type of dictionary in progress. The definition’s wording is constrained by column-inches, the Longman defining vocabulary, and the permissible range of syntactic structures. Splitting, by contrast, is only marginally constrained, by the dictionary’s positions on lumping *vs.* splitting and on subsenses.

Inclusion is generally the most extensively reviewed aspect of a dictionary in the non-specialist press, yet was not deemed ‘hard’ by the LDOCE3 lexicographers. This may be

The hard parts of lexicography

Could you please rank the aspects of lexicography listed below, putting 1 beside the hardest, 2 beside the next hardest, etc.

Some things will be hard and will arise quite often, others, as hard, will arise less often. Please put, in the second column, '1' for very often, '2' for quite often, '3' for rare.

Please add any hard aspects of the work which you feel I have left out at the bottom. Also please add comments expanding on your answer where you have more to say.

YOUR NAME:

	How HARD	How OFTEN
1. Inclusion, headwords		
2. Inclusion, senses		
3. Splitting; identifying the senses for a word		
4. Finding right wording for definition		
5. Knowing when you have spent long enough looking at the corpus		
6. Multi word items: deciding whether to show them at all, as lexunits or collocations, how to show variability		
7. Priority of different aspects of meaning and use in the definition		
8. Examples - inventing/selecting		
9. Extent: how long to make entries		
10. Consistency with definitions of similar words elsewhere in the dictionary		
11. Inflections, variant forms		
12. Grammar		
13. Run-ons		

Figure 1: The questionnaire

because the task of searching for new words was not undertaken alongside other lexicographic work, so the subjects did not view it as part of core, everyday lexicography; or because, whatever the generalist press may say, inclusion is not critical for a monolingual, advanced learners' dictionary; or because a decision could be based on a combination of corpus frequency and whether a number of previous editions and competitors included it; or simply because an inclusion decision is a yes/no decision which, once made, requires no further thought.

It is noteworthy that the matters often requiring the most detailed specifications in Style Guides, and which frequently seem the hardest to grasp to users, do not cause lexicographers undue strain. LDOCE3 was written on an SGML-based editor which only permitted appropriate kinds of information in appropriate fields, and the Style Guide included very detailed specifications for grammar, inflections and variant forms. Lexicographers became very good at following the rules, and did so without great difficulty. The only two comments made in the 'any other comments' section of the questionnaire were that these tasks were not intrinsically difficult but became difficult where the Style Guide was not clear enough.

Perhaps the most significant message of the survey is that the hard tasks are those which it is not straightforward to teach or give rules for. The hardest task is choosing the wording.¹ This tallies with a comment from Michael Rundell, Managing Editor of LDOCE3, that "one of the first things I look for in a lexicographer is someone who can write well". Choosing words well is closer to poetry than linguistics, and is not easily taught.

The second hardest part is splitting. This is a matter on which textbooks have nothing to tell us, and lexicographers' training is entirely example-based, largely because rationales for lumping and splitting are so little understood, even by lexicographers who do the job well (Kilgarriff 1992).

In conclusion: this was a simple survey. It is unclear to what extent the findings would apply to native-speaker, or to bilingual lexicography. Yet it does make some points about the nature of lexicography and the challenges to be overcome in lexicographer training.

Reference

- Kilgarriff, A.** 1992. *Polysemy*. DPhil thesis, University of Sussex.
Longman Dictionary of Contemporary English. Third Edition. 1995. Edited by Della Summers. Longman, Harlow.

¹This would not apply to bilingual lexicography, where defining is translating.