Breaking new ground

A review of
Cambridge International Dictionary of English, *CUP, 1995*

Péter A. Lázár

0 Introduction

The average language learner regards the dictionary as a source to check meanings, spellings, and not much else. This review, written partly from the perspective of the language teacher and the "applied syntactician", recognizes this; also, it assumes a general readership rather than one trained either in lexicography or syntax. I don't suppose, for one thing, that in matters of dictionary design and use there is agreement on the area where the lexicon shades into the grammar, or vice versa. Also, I doubt that a review of a dictionary for students of English is worth its while if it merely addresses a few dozen academics working in these fields.

Nothing that is remarkable about a dictionary to a linguist is intrinsically attractive to a language teacher; no design feature of a dictionary which a learner finds useful or novel is guaranteed to meet the approval of the teacher, and even less so of those professionally involved in matters lexicographic. One obvious fact illustrates this fairly well. Although both applied linguists and lexicographers attach great importance to the authenticity of the language used in dictionaries, usually taking pains to get style, register, medium and context right, it is often the teacher who is the only agent in the language-learning/teaching enterprise who knows that most learners just don't believe you when you say that a dictionary is useful for *language learning* as well. If dictionaries have so many faces;[1] if editors, users, and teachers have so diverse perspectives (even if they may share common interests); and if the dictionary which teachers value highly in their daily classroom work is not necessarily the same that they find most useful as an aid to marking students' papers, then a conscientious review will have to be either one that singles out just one of these perspectives and takes a detailed look at it, or one that catalogues almost all these points, offering a shorter, but perhaps balanced description of them. I have opted for this second task, of taking many focuses, concentrating thereby on the

applied linguist, the language teacher, and the user/learner[2] at the same time.

Nothing or almost nothing will be said about style labels, national varieties, selection, collocation, and various aspects of the lexical–grammatical information presented. These will only be touched upon to the extent that they are relevant to the three areas which are central here, being looked at from several angles: authenticity, senses, and definition. Very little will be said about the *False friends*, which, although one of the indisputable welcome features of CIDE, are essentially peripheral, add-ons which could be removed without affecting the rest of its structure.

1 What CIDE is claimed to be

Each of the freshly published dictionaries as well as each of their new editions is in the convenient position of being able to learn from its own as well its predecessors' mistakes. Ideally, then, any new venture ought to unite all positive features of all previous dictionaries, while avoiding the pitfalls of all earlier works. If indeed this were how it works, full perfection would have been achieved a long time ago.

Most new works, however, in addition to the task of synthetizing, decide to break with the traditions and offer something radically new, often *instead* of the time-honoured ways.[3] We tend to forget that there was even a time when Verb Patterns were revolutionary.[4] The idea of a controlled defining vocabulary has been with us for quite long (Landau 1984 : 13–15),[5] but it only found its way into EFL dictionaries with the advent of the first LDCE. Detailed usage information, self-contained language notes, a separate column for lexical relations (basically, synonyms and antonyms) were unheard-of in the first edition of the OALD, the model for all later EFL dictionaries.

Consequently, if a publishing house — the world's oldest — which has never published a comparable dictionary risks launching one incorporating the best of the lexicographic tradition but also introducing quite a few novelties, then this must be a thoroughly motivated move and ideally, a work that is designed and written well. Competition seems to get tougher and tougher in this business. Legend has it that Cambridge University Press declined to publish what was to become the OED because James A. H. Murray, the chief editor of the OED, asked for £100, which the then CUP thought too much. Now CUP boasts Paul Procter, one-time chief editor of the well-known and well-proven Longman dictionary series. Procter's undertaking means business competition for three dictionaries or

rather whole traditions—including his own—at the same time: OUP (the "Hornby" dictionaries), Collins (the Collins Cobuild dictionaries) and the Longman dictionaries. Although ours is certainly not a comparative evaluation of these four "lines" of dictionaries, there are areas where ignoring their existence would have been imprudent and comparison seemed especially instructive. Some quantitative characteristics, for example, virtually offer themselves for the analysis of such correlations.

1.1 An overview of the promotion materials for CIDE

Most promotion materials emphasize that CIDE is the first dictionary from CUP for learners and users of English as a foreign language from intermediate level upwards, to which they actually refer to as a "landmark in our publishing development." The *New From Cambridge* leaflet (NFC 1995) mentions the most important features of CIDE under three headings; the key words are as follows:

- **Comprehensive:** 100 000 words and phrases are arranged under 50 000 headwords; more than 100 000 example sentences demonstrate usage and context; more than 2 000 items are illustrated
- **Clear:** Guide Words differentiate immediately between sense of the same word; a *Phrase Index* gives instant access to 30 000 phrases and idioms;[6] there are uncomplicated grammar codes, to which always an example sentence is attached
- **International:** it offers a thorough treatment of differences between British, American and Australian English; it contains *False Friend* information for sixteen languages.

Another apparently innovative feature is the "extensive research among thousands of teachers and learners of different nationalities" that preceded the compilation of CIDE. According to the promotion materials, this has made possible the clarity of layout and explanations, detailed but simple grammatical information, and a user-friendly dictionary in general. Also, NFC 1995 explicitly states that CIDE has more example sentences than any other (learner) dictionary.

The three innovations that all advertising materials point out are the Guide Words (see **4.2**), the Phrase Index (see **5.2**), and the False Friend information (see **4.3**). The (grammar, usage, vocabulary and style oriented) Language Portraits are not singled out as innovative, although bringing together these under the same heading — a collection of mixed content,

including titles like Adjectives, Age, Apostrophe, Australian spelling, Borrow, and Britain—is unusual indeed.

1.2 Why "International"?

The word "international" may strike one as somewhat enigmatic in the subtitle of a dictionary. The clues as to what it signifies, however, are ultimately not hard to find. The (undoubtedly novel) False Friends information is generally hailed as one of CIDE's "international" features. Another facet to internationalness is that the editors' aim has been to "make the cultural content of CIDE as international as possible, reflecting the fact that English is often the only common tongue between groups of speakers of other languages." This is a doubtless laudable aim, but I cannot imagine how it works and in what type of entries; more importantly, any drive to do this is likely to work against a true presentation of the culture-specific aspects of a language.

Another — I should think, the genuine — sense in which CIDE is "international" is that it explicitly targets users—learners and teachers of English — of all nationalities. I suspect that "International" may simply be a more polished word for EFL/ESL dictionary, the real message being that CIDE is an (advanced) learner's dictionary.[7] There is no reason to believe that CIDE will be less successful there than any of its competitors, but whether it proves more successful worldwide will finally depend not so much on the declared "international" components but the entirety of its lexicographic merits.

1.3 Chief editor's Foreword

Clarity and simplicity, "the clearest presentation [...] with the minimum of the fuss and clutter that are the usual feature of dictionaries," avoidance of "cumbersome numbers": these have been singled out as the main concerns of the CIDE editorial team. We also learn that "each entry is for one core meaning," to which the reader is immediately directed by the Guide Words. The *Foreword* briefly informs about the software resource background used and how it augments the possibilities of the EFL component. The Cambridge Language Survey is mentioned here (while page xii provides a glimpse of the structure of its collection of over 100 million words, broken down by variety/source type). Speed of access is highlighted as an element of clarity, and attention is brought to the extra material: the Language Portraits and the pictorial illustrations. Completeness in terms of number of entries and examples as well as of non-British usage, new words, cultural

content, even sensitivity and impartiality in the treatment of gender, race and religion are also mentioned.

What exactly the unmistakable features of CIDE are and how readers can make the most of them is explained in the *How to...* passages, which we will briefly survey in the next section.

2 What CIDE has to say to the user

The editors of CIDE have acted upon the realization that the average user can rarely be bothered to study lengthy introductions, prefaces, even detailed guides to their dictionary. Accordingly, two things strike the reader about the front matter of CIDE: size and style. It must be a deliberate break away from the custom in many earlier dictionaries—a most welcome move—that both the Foreword and the *How to...* sections as well as the Grammar: The Parts of Speech pages have been truly written for the reader and not for the profession. Some of the examples: "A single word in bold shows that it is often found with the word being looked up"; "When a single word has more than one meaning, Guide Words help you to find which meaning you want"; "Some words are not given in their alphabetical place because the different parts of speech of a word are grouped together when they share a similar form and meaning." No cryptic talk here about collocates, selectional restrictions, lemmata, run-ons, homographs, homonyms, derivation, compounds, inflections and similar abstruse things.

The *How to find words and meanings* section is a no-frills guide to single words and words in groups; the two-page *How to use the dictionary* introduces the rest of the machinery, while the Grammar section contains Part of Speech information and refers the reader to the relevant Language Portraits.

3 General

3.1 Format, layout, entry design

Sold in hardback, flexicover and paperback binding, the 23.5 × 15.5 cm volume is 1 792 pages, which makes it an ideally sized desk dictionary with clear and simple page design, black-and-white illustrations—an impressive, "serious" book for the exacting advanced learner–user.

The inside front cover lists the grammar labels and the usage labels and abbreviations in the dictionary. The rest of the front matter contains a page with the Editorial Team and Consultants, one with the English Language Teaching Consultants and Academic Consultants, another listing the

Subject Advisers, followed by the Contents, a Foreword, and the sections How to find words and meanings (one page), How to use the dictionary (two pages), The Cambridge Language Survey, and Grammar: The Parts of Speech (five and a half pages). The back matter consists of the Defining Vocabulary, The Phrase Index, and The Pictures, Language Portraits and lists of False Friends. There are no independent lists for the latter three, and this may make the job of their separation somewhat difficult: it must be said that the mixed nature of items may be disturbing, and the asterisk put after those that are Language Portraits does not make them stand apart from the Pictures.

The information in the Language Portraits, these large panels in the main body of the CIDE, is the most heterogeneous, comprising culture, grammar, pronunciation, punctuation, spelling, usage, vocabulary and style. Under the letter H at the back we find, among others, *Holidays; Homophones and homographs; Hundred* and *Hyphen,* while the T section contains such diverse things as *Telephone; Tenses; There; Time; Titles and forms of addresses.* Since the Language Portraits are so mixed themselves, it is regrettable that LP's, Pictures, and False Friends are not listed separately.

The Language Portraits are fully cross-referenced indeed, but sometimes the titles themselves do not match their content: the LP⟩ *Continuous form,* for example, which has only one small paragraph about the continuous tense, refers the reader to the LP⟩ *Tenses,* and discusses the *non*-use of the progressive.

Those LP's that centre on culture/vocabulary add to CIDE the most useful ingredients of a thesaurus: *Borrow; Crimes and criminals; DO: verbs meaning 'perform'; Expensive;* and *Eye and seeing* are cases in point.

The entry design in broad sense, which is bound up in complex ways with the new feature of the Guide Words and with questions of defining, shows quite a few innovations (see **4.1, 4.2,** and **5.2**). Supported by a clear, "pure" page layout and the sparingly but uniformly used three typographical devices of bold face, italics, small capitals, it undeniably enhances the speed of lookup. As the promotional brochure, NFC 1995, informs us, the "extensive research among thousands of teachers and learners of different nationalities" showed that what users want most is clarity of definitions, descriptions and examples, which the editors have achieved through a "generous text area" and "unfussy layout." The font size is just about ideal, and so are all aspects of visual/graphic presentation including the pictures. The running heads on the outsides of the pages repeat the first and the

last entry of the page in the usual fashion (**baulk** to **be**; **be** to **bear**), while pagination comes at the centre, with "page" spelt out in full.

A conspicuous feature of the two-column pages—one whose function the casual reader will not immediately realize—is the smallish line numbers running down between the two columns. These are in rather small print (as are the page numbers), but perhaps bigger size would be too heavy and disturbing, especially if one does not make use of the numbers (see **5.2**).

It may be interesting to note that the longest entry is **go**, taking up five whole pages and containing over fifty senses. Second place goes to **get** and **run** (39 meanings each), **do** is third (36 meanings), **make** is fourth (32), **out** and **on** take 5th place (31 meanings each), followed by **take** and **come** (sixth place: 30–30), **put** (27), **set** (25), **up** (24) and **in** (21) (CUPPR 1995).

This may be the best place for a critical remark too. A problem that has to do with the entering of material, which may seem a minor one at first but turns out to crop up in many places, can be very annoying. Many of the *Language Portraits, False Friend* panels, and the illustrations are extremely hard to locate, because they are not entered where they alphabetically belong. The *Czech False Friends* box is on the **cud** to **cue** page, which is not justified either by *Cz* for *Czech* or *čs* for the name of the country (see also **3.3**).

Whether one thinks these codes are a good way of referring to countries may be a matter of taste; I, for one, would never dream of searching *Dutch False Friends* on page 957 because of NL. Alphabetically, they ought to come between **nix** and **no**, a solution which is technically impossible, so they wedge themselves between two portions of the entry **no**. On page 66, the entry **article** GRAMMAR sends you to the LP⟩ *Articles* providing no page number, and the LP⟩ is just two pages away, titled *The definite and indefinite articles, 'the', 'an' and 'a'.* The layout reasons for this may be clear (there is not enough space for the panel on the same page), but the fact is still annoying. Page 1304 contains **sexism** and refers on to the LP⟩ *Sexist language*, which is on the opposite page but its title this time is *Using language that is not sexist.* On page 551, the entry **form** SHAPE refers you to the LP⟩ *Forms of words (spelling)*, which is right on the next page, but the title is neither *Forms...* nor *Spelling...* but *Word forms: spelling rules.* If you have used this panel once and want to find it again, it is simply unguessable whether the LP⟩ is at the letter F, at S, or at the letter W.

We have seen that none of the tables usually found at the back of dictionaries (except the Defining Vocabulary, Phrase Index, and the "contents page" of Pictures, Language Portraits and False Friends) is included

in the CIDE back matter.[8] One wonders whether all the difficulties just mentioned could not be removed if all such stuff (apart from the pictorial illustrations for words, of course) was accommodated in tables at the back, using proper alphabetical order, no repetition of titles in different forms. I may be wrong, but I suppose much of this kind of information is still thought of as "the Appendices": Abbreviations, Spelling, Money, Punctuation, Time, Word formation, Prefixes and suffixes ("word beginnings" and "word endings"), Letter writing, Mathematics/Numbers, Weights and measurements, Nations and nationalities/Geographical names (including those of England, the USA *etc.*), Irregular verbs, Opposites, Military ranks, Forenames, Pronunciation/Transcription, Titles and forms of address, even Family relationships, Calendar, Chemical elements, Animal Table, Books of the Bible, and Works of Shakespeare.[9] Some of these have always fared better: common abbreviations are usually there, but in the body text; irregular verbs have survived, just are not listed separately. Some, however, seem to be gone forever: the useful Common Forenames feature or the not-so-useful Chemical elements are gone along with the Bible and the Bard.

Aware as we are that the dictionary–encyclopedia boundary is one that may or may not exist, let us look at CIDE's encyclopedic aspects. One of the more clearly encyclopedic features, the thesaurus character of CIDE, is enhanced by some Language Portraits and the Pictures to which latter the reader is referred by the PIC flags. Because today's EFL dictionaries in general offer a wealth of diverse information (both knowledge of language and knowledge of the world, both verbally and pictorially presented, both in alphabetical and thematic arrangement), and because the separation of linguistic information itself into grammatical and lexical is far from easy, one cannot marvel at the Language Portraits being quite heterogeneous. Inclusion of information on synonyms already means parting from the alphabetical order, while the pictorial medium is intricately bound up with encyclopedias. And when Language Portraits like *Opposites* catalogue how negative meanings are expressed, or when pictures such as *Neck* list **jumper**, **person**, **bottle**, **violin** and **guitar**, they mould all of that into something necessarily eclectic. In a carefully informal way, we could perhaps talk about not one dictionary–encyclopedia contrast but rather a dictionary–thesaurus–encyclopedia continuum along which all of the EFL dictionaries under discussion are continuously moving.

There is a noticeable shift in the development of all EFL dictionaries that as they reach their next edition, rare words tend to become smaller in number, for this is the only compromise that can possibly be made to ensure that what are judged to be the new words that promise to be lasting

as well as the new kinds of information can be included. Coverage issues are intricately linked with up-do-dateness, authenticity, and inclusion or exclusion of material. It is fortunate that the desirable tendency towards the everyday, the colloquial and the natural does not necessarily counteract the striving for accuracy, which is a requirement in the more "encyclopedic" definitions even if the *style* used in such definitions is generally not appropriate for a learner's dictionary. We now learn from the *New From Cambridge* brochure (NFC 1995) that in specialist subject areas such as law, business and medicine the entries have been checked for factual accuracy by experts.

Let us illustrate why this is important, and what healthy compromise means in the CIDE definition styles. The definition of **cursor** in CIDE is 'a movable marker on a computer screen which shows the point where the work is being done' is a good one. The definition of Hungarian *kurzor* in Bakos 1994 has been criticized — quite rightly — for its inaccuracy: 'villogó jel a számítógép kijelzőjén, mely a beírandó karakter helyét jelöli [blinking sign on a computer's display marking the place of the character to be entered].' This may be the average PC user's first approximation, but it certainly does not qualify as a good definition.[10] With an EFL dictionary, it may be debated in the case of each and every entry whether encyclopedic rigour or truthfulness to the ordinary person's common sense should have priority, but it is reassuring to know that wherever relevant, CIDE has tapped into specialist expertise and seen to it that what is "scientific" precision is maintained in otherwise readable definitions.

Up-do-dateness and authenticity will be discussed in **3.3** and **3.4,** while the technical aspects of the organization of entries — their internal structure, assigning headword status, demoting and promoting items — will be discussed in section **4.1.**

3.2 Coverage

While in the case of LDCE, OALD, even CCELD, one can determine the changes from each previous edition to each new one, here with CIDE it is like starting a clean slate. While we have to do with a learner dictionary that is ideally the three mentioned above, yet there is no work of really comparable character against which CIDE could be measured. Useful comparisons and relative statements concerning numbers of headwords, entries, definitions, *etc.* are almost impossible to make anyway. The table below, which contains numerical information gathered together from the relevant dictionaries, will amply illustrate this.

Set off by double line are the most recent two EFL dictionaries, OALD 5th edition and CIDE itself. Where no figure is given but the count is easy, I have included the number, in italics.

	Parameters	OALD 2nd ed 1974	LDCE 1st ed 1978	LDCE 2nd ed 1987	CCELD 1st ed 1987[a]	OALD 5th ed 1995	CIDE 1st ed 1995
A	Citation corpus, words	—	—	27 mn	20–400 mn[b]	100+40 mn[c]	100+ mn
	Defining vocab., words		2 000	2 000	?[d]	3 500	2 000
	Culture pages					8	
	Usage notes			400		220	
	Language notes			20			
	Language study pages					16	
	Language portraits						100+
	Appendices	*10*	*7*	*3*	—	*10*	*2*[e]
	Illustrations	1 000		500	—	1 800	2 000
	Items with phonetics	100 000					
B	Headwords						50 000
	Headwords & derivatives	50 000					
	References				70 000	63 000	
	Meanings of ws & phrases			83 000			
	Definitions					65 000	
	Words and phrases			56 000			100 000
C	Example sentences						100 000
	Examples			75 000	90 000	90 000	
	Illustrative phrases & Ss	50 000					
D	Idioms and phrasal verbs					11 600	
	Idiomatic expressions	11 000					
	Idioms/phrases						30 000

[a] 2nd impression

[b] 7.3 million word corpus; total computerized corpus of over 20 mn words; running length of 400 mn words (20 mn × 20-word context; Carter 1989 : 34)

[c] 40-million-word Oxford American English corpus

[d] CCELD mentions a "carefully controlled defining vocabulary" but gives no figure.

[e] The *Phrase Index* and the *Phonetic Symbols*—strictly speaking, not Appendices.

There is not much to be said about section **A** of the table. Computerized citation corpora were obviously not yet in in 1974, and the idea of a pre-determined defining vocabulary based on frequency list was conceived. The following lines—from *Culture pages* to *Appendices*—provide

information about various types of appendix-like extra material; *Illustrations* always refers to pictorial illustrations; *Items with phonetics, i.e.,* the number of transcribed words, is only singled out in OALD.

It is the three sections **B, C** and **D** that are very difficult to compare. In **B,** the CIDE wording is the following: "100 000 Words and phrases arranged under 50 000 headwords," and it is not clear how the OALD figure can be so small, given that it includes derivatives as well. LDCE's "Words and phrases" figure probably does not cover the same thing as the corresponding CIDE figure, although the difference in size (LDCE is 1229, while CIDE 1701 pages) may warrant the difference. The signification of "References", "Meanings of words and phrases", and "Definitions" is even harder to guess. These terms reveal very little about the exact number of lexical items, of entries and subentries, about the items properly defined in a dictionary (*cf.* Landau 1984: 84–88).

Coming to **C,** we see that what counts as a full example sentence is another delicate question. For the data in this section to be reliable we would have to know, *e.g.,* in the case of OALD, what percentage of the examples are phrases, and how many are actually sentences. Whether phrases or sentences, we can probably speak about a steady increase in illustrative material.

In **D** it is apparent that although only CCELD separates idioms and phrasal verbs (the OALD figure may not include them), CIDE's 30 000 does cover all kinds of multi-word verbs as well, which heavily distorts the picture in this section again (on multi-word verbs in CIDE see sections **4.2** and **5.2**).

Another approach to the quantitative evaluation of a dictionary is when one multiplies the number of lines to the page by the number of letters to the line. In the case of CIDE this yields: 84 lines × an average of 55 letters × 2 columns = appr. **9 240** characters. To see what this means, let us do the same calculation for CCELD this time, which has appr. 85 lines × appr. 50 letters × 2 columns, *i.e.,* **8 500.** While CIDE, whose top and bottom margins are smaller, has exactly 84 lines on every page, it can print ten percent more in a line because the spacing between columns is not used up by the *Special Column* notes, a special grammatical/lexical feature of CCELD. Moreover, because of the narrower margins it can use a type which is much more readable than CCELD. Since both dictionaries are 1 700 pages from A to Z, CIDE proves slightly superior in terms of the information presented: appr. 15 700 000 against CCELD's 14 450 000 — an edge of about one million *more readable* letters.

3.3 Up-to-dateness

It may be difficult to establish what those features are that give a dictionary a feeling of up-to-dateness in the eye of the general reader. Factors such as layout, typeface and illustrations are likely to play as great a role as a generous but selective inclusion of all those new words that the native speaker may specifically try and look for in a new dictionary. Unlike the native speaker, the average foreign user rarely realizes that they may need a new edition of their favourite dictionary every now and then; first, if they feel they have to keep abreast of changes, it will be because of the technical terminology (this mostly means computers and information in general these days), and second, because they know that slang is dangerously ephemeral. As has been pointed out in **3.1** above, CIDE is acutely aware, and acts upon the realization, of the importance of the *former*. In **5.3** we will see an example of how attentive CIDE is to the needs of readers as far as the *latter* is concerned, doing an especially good job of presenting the vernacular of the four-letter kind in all its pragmatic versatility.

Perhaps the significance of the up-to-dateness of a dictionary simply in terms of new coinages ought not to be exaggerated. All dictionaries will have a list of words which their editors proudly announce as new inclusions. The following list contains some of the words which, according to various promotion sources (Bookseller 1995, CUPPR 1995, Procter 1995), have made their dictionary debut in CIDE:

> **aggro, bumbag, chatline, chocaholic, cred, DINKY, diss, docudrama, dork, dosh, Exocet, fanny pack, fanzine, feel-good, filofax, flexitime, 4WD, grunge, jobshare, kissa-gram, laddish, liposuction, lippy, lummox, Lycra, medal-lion man, mini-series, minipill, MRE, multiplex, NIMBY, oik, ovenable, veggieburger, prequel, PWA, Semtex, sky-jack, slasher, snit, stonewashed, televangelism, trophy wife, veggieburger, wannabee, wonk, wuss, zilch**

Given the limitations of space, there is always a legitimate excuse for incompleteness in any, but especially in a learner's, dictionary. Any hit-or-miss comparison is bound to be haphazard and inconsistent. The new OALD 95 now includes **wysiwig**; so does CIDE. **Gendarme** and **gendarmerie** aren't in CIDE; LDCE 1987 has the former but not the derived noun. The liquid **must** — certainly not among the new coinages — is not in CIDE; it is included in all other works I have checked, and though we have no proof that it has been negligently left out, it is not one of those

marginal items that must go first when the newcomers demand space. (The same **must**, incidentally, is also missing from CCELD, and I cannot help the thought that both have simply been lost—having fallen prey to a kind of entry organization which has too many separate (sub)entries rather that two or three well-delineated numbered senses).

Coverage and up-to-dateness belong together in obvious ways. We have seen in **3.1** that much of the "marginal" material, such as the False Friend tables, is difficult to find in CIDE. Incidentally, there is no entry for the word *Czech* itself, only for **Czechoslovakia**. I wanted to check whether there is a *Countries* panel so I could ascertain if there is an everyday, shorter word for what I usually refer to as the *Czech Republic;* the answer is *no* to both questions.

There is another specific respect in which CIDE is unquestionably up-to-date: its treatment — both in the entries and the Language Portraits—of all those words especially dangerous for the foreign learner which may be offensive because of their sexual, racial or religious content, ranging from the well-know taboo items that were marked as such in most dictionaries twenty years ago, to the most recent PC lingo. The approach is fundamentally unbiased and descriptive, although the two emphatic titles of the two opposing columns in the Language Portrait *(Using) Sexist language*—"Old-fashioned sexist language" and "Modern non-sexist language"—may seem a bit of an exercise in *categorism.* The entry for **man**, then, in which the reader is referred to this Language Portrait, compensates for this by warning the zealots that not all that glitters is gold: **manager** and **manufacture**, alas, have nothing to do with 'man'. Interestingly enough, **bitch** 'unpleasant woman' and **cunt** 'very unpleasant or stupid person' (used of men as well!) are not marked as "sexist".

I find it unfortunate that the individual entries do not consistently indicate words as being "sexist": the entry for **businessman** does not refer to the LP *Sexist language* (where readers are advised to use **executive**); both **cameraman** and its non-sexist version **camera operator** feature within the entry, but no indication of sexism here either; **milkman** (ironically, "the *person* who delivers milk...") and **mailman** contain neither alternatives nor a warning; **postman** gives (the allegedly only-American) **letter carrier** but not as a non-sexist variant; **chairperson**, **chair** and **chairman** share the same entry (in that order), but here, too, there is no reference to the Language Portrait *Using Sexist language.* **Gasman** has no LP flag either, and its meaning is supposedly "a *man* whose job is reading METERS..." Though not difficult to find (both would be between **layette** and **lazy** anyway), **layman**—both in the 'outsider' and the 'religion' sense—is hidden within

the entries for **layperson** UNTRAINED and for **layperson** CHURCH. I even tried **poetess**, but it is not in; there is no such entry: even **poet** is run in within **poem**, along with **poetry**.[11]

When all of these entries fail to call the user's attention to potential dangers, then the case of **fireman** is simply unpardonable. There is no such entry[12] either: this word is only given a look-in as a second option within the entry **firefighter**, suggesting that **fireman** may be following **poetess** on its way out of modern non-sexist language; but then, strangely, there is no hint as to sexist usage in the entry.

3.4 Authenticity: its basis and manifestations

Authenticity as a forte of CIDE in general is not as foregrounded as in the case of the Collins Cobuild dictionary, which was hailed as a dictionary that "helped learners with real English"[13] (quite unfairly suggesting, I think, that the others might not be). When the work of lexicographers first involved computer databases, these barely contained more than one million words (hardly enough to keep tabs on slang, a variety of registers and geographical varieties). The vast amount of language in CIDE has been selected on the basis of a 100-million-word database, which the promotion materials say is the largest that has ever been put to work for dictionary making. The basis on which the many years' language research and analysis that had gone into CIDE was done is called an "integrated language database."

This huge computer resource — the Cambridge Language Survey (CLS) — , of which Paul Procter is the founder and director, combines the 100-million-word English corpus with a dictionary compilation system. The sheer size of a computer database hardly guarantees a good dictionary, and even less so a good learner's dictionary. Here, however, more than this is involved: a non-native speaker corpus, collected systematically from exam scripts by examining bodies in different countries.[14] Much of this information surfaces in the False Friends (CIDE 1955 : viii). Hungarian, which is unfortunately not included, seems to be one of the seven languages in which the meaning of the "local" word for the Hungarian *aktuális* is not 'real, factual, existing', but 'topical, timely, current'. This is indicated by the list of nationalities at the bottom of the CIDE entry for **actual**, referring the reader to the False Friend section for ČS, D, DK, E, F, PL and RUS. (I am sure that this kind of information will be valued as much by conscious speakers of these languages as by serious browsers of all nationalities).

Authenticity, of course, here as well as in other dictionaries, shows up virtually everywhere: in the selection of the entries, in the breakdown of meanings, the provision of collocations, idioms and example sentences, in the inclusion of the most recent (but hopefully not ephemeral) lexical items, even in the pictorial illustrations where, too, objects become outdated quicker than many people think.

CIDE gives no clues to the ways in which the words and expressions were culled from the vast database of CLS (it only mentions the software tool); nor does it discuss the principles of processing and selection. We do not learn either from the promotion materials or from the dictionary itself whether all the examples given appear in their original, unedited form, or are edited/adapted, or whether some of them are ultimately made-up, or possibly that some clever combination of these has been used, depending on didactic aims in general or individual entries in particular. Maybe some people think you do not have to make up examples these days: after all, when you have a huge corpus as this you can hardly hope to be as imaginative as to invent more than what comes your way anyhow. On the other hand, colourless examples whose authenticity is only manifest, say, in that they contain proper names help very little. If the corpus yields no better for a given word in a particular sense and a specific pattern, then they must be tampered with. There is not much point in being 100 per cent authentic in terms of examples (not restricted by any didactic considerations), and at the same time using a carefully selected (and thus necessarily limited) defining lexis—unless, of course, the database is so vast that it even allows this: to only use those examples out of the practically infinite number that only contain your "survival" vocabulary. On the other hand, authenticity can also backfire, as proven by many of the clumsy and often criticized Cobuild sentences (Hausmann & Gorbahn 1989 : 46); a host of true-to-life sentences which fail to illuminate a sense, just show off; and sentences so heavily filled with authentic but esoteric or idiosyncratic contexts that they require a better-than-average imagination, detracting from the definition at hand.

I think that if you are to have didactically successful examples that not only illustrate but also support the definition, then a measure of editing and adapting is needed—and I feel that CIDE's examples are exactly like that: the results of a competent and careful process of adaptation.

Although promotional materials do not boast about this, the importance of the inclusion of well-known quotes from popular songs, television, films and books would be hard to overemphasize. This is not restricted to culture-specific entries. The word **cynic** is illustrated by a quotation from

Oscar Wilde, more succinct and elegant than many definitions: *"What's a cynic? A man who knows the price of everything and the value of nothing."* This is useful and refreshing, and it smuggles in the best features offered by dictionaries of quotations or, *e.g.,* Longman's *Dictionary of English Language and Culture* (LDELC 1992).

One sometimes wonders, however, whether some of the quotations really *elucidate* the meaning of a particular word. The word **beam** [LIGHT], for instance, rather than "come alive" in the expression *beam me up, Scottie* (from Star Trek; Procter 1995 : 2), further mystifies—and not only because the reader is ignorant of popular culture (*cf.* **4.1**). Other illustrations, however funny they might appear, are of no real didactic value: the information in the entry of **back** that *"I'll be back"* is 'an expression used by Arnold Schwarzenegger in the Terminator films' has nothing *linguistically* relevant, and is not particularly hilarious either. If you are one of those beginners who have seen the film (in English!), it may help you not to forget an "expression", otherwise it is not too helpful.

4 Words and meanings

4.1 Entries, senses, definitions

The most important qualities of a dictionary are revealed by the way the entries are broken down, senses grouped and defined. In this longest section the word **beam** as presented in CIDE will be scrutinized, and the new approach taken to the design of entries will be illustrated.

There are three entries for **beam**, each with a Guide Word; the first is both a noun and a verb, the second, only a noun, while the third, only a verb; the entries in skeletal form (omitting transcription and using the ellipsis sign to replace the irrelevant parts of the examples) are repeated here as follows:

beam [LIGHT] *n* [C] a line of light... • ... *in the weak beam of the torch* • ... *mesmerised by the beam* **of** *the car's headlights* • ... *the* **laser** *beam* • ... *syringes using an* **electron** *beam* • See also MOONBEAM; SUNBEAM
beam (*obj*) *v* • *the ... sun beamed* [=shone brightly] **down on** *the boat...* [I] • *The concert was beamed* [=broadcast] *by satellite...* [T] • (*fig.*) *This mailing technique is used to beam* [=direct] ... *services to likely consumers* [T] • *"Beam me up, Scottie"* (popular phrase based on the Star Trek television series)

beam WOOD *n* [C] a long, thick piece of wood, metal or concrete... • *The sitting room had* **exposed wooden** *beams* • In the sport of women's GYMNASTICS, **the** beam is a wooden bar on which the competitors balance... • (*Br infml dated*) If you are **on** your **beam-ends** you have little or no money left. • PIC Sports

beam SMILE *v* to smile with obvious pleasure • *She beamed with delight...* [I] *The child beamed at his teacher...* [I] *"I'm so pleased to see you," he beamed* (=said as he smiled). [+clause]

Guide Words will be discussed in section **4.2.** Here, a brief comparison of the **beam** entries in OALD, CCELD and LDCE will serve as a starting point in our attempt at highlighting some of the issues related to lemmatization, entry layout, and definition.

OALD has just one **beam** entry (which in most dictionaries would suggest polysemy), containing a noun **beam** with the meaning 'wood' and another meaning 'ray', followed by the verb **beam**. In the CCELD, **beam** is also one entry, consisting of nine numbered paragraphs, in which nominal and verbal meanings follow each other in a mixed order (*v, n, v, n, v n, n, n, n*). The LDCE provides three entries, **beam**^{1} *n* 'wood', **beam**^{2} *n* [both 'light' and 'smile'] and **beam**^{3} *v* [both in the 'light' and 'smile' senses]. Thus, while in OALD it is simply difficult to see how the senses of **beam** are structured, CCELD gives a rather distorted picture of this structure and LDCE is accurate but perhaps too detailed and not very friendly, CIDE strikes a balance between simplicity of exposition and precision of analysis.

CIDE's use of the *obj* label, both with and without parentheses, however, is very awkward. The label (*obj*) suggests not that the verb which it follows is *either* transitive *or* intransitive; it rather looks to be *parenthetical* information about the verb being *transitive*. (Incidentally, *Grammar labels in the dictionary* does not list the label (*obj*), which only appears in the *How to use the dictionary* section).[15] One could also argue that the label *obj* should not precede all of the transcription, sometimes the Guide Word, the part of speech symbol and subcategorization label [T], but follow them. In this way the first thing CIDE teaches about one of the **bear** words, for example, is that it is (*obj*); this is first followed by the Guide Word box ACCEPT, then the transcription; only after that does one learn that this is the verb **bear**. The other problem here is that it is not prudent to refer to transitivity in two different ways, using both *obj* and [T] *vs.* [I]. Readers, I suspect, are used to these two letters, which suggests that something like [I/T] or [I or T] or [I; T] (in this or the reverse order) would be a better notation. Whether the separate symbols [T] and

[I] should *follow* the example sentence (as they do in CIDE, presumably for reasons of clarity and readability) is questionable: after all, *after* you have read the example, you will know whether the verb is transitive and you hardly need a letter T.

It will be apparent that the *verb* in the **beam** LIGHT subentry is not defined properly, or at least not in the same fashion as the *noun* **beam** LIGHT. This is because the function of the • signs is far from obvious: they do not always separate *different senses* within (sub)entries; indented subentries, in turn, may not always contain *derivatives* of the headword;[16,17] finally, at the level of entries, there is no guarantee that all words with a Guide Word of their own are *homophonous*.[18] Repeating this in a reversed order: if there is no uniform key as to what kinds of words are assigned to a separate entry (possibly with a Guide Word); if there is no uniformity about what counts as a run-in subentry; and also, if there is no safe clue concerning the • sign, then this probably means that everything has been subordinated to practical considerations.

What the *How to find words and meanings* section comments about these three things is as follows:

- (on Guide Words:) "When a single word has more than one meaning, Guide Words help you to find which meaning you want"; and "... with some words the different meanings shown by the guide words are not [...] clearly separate [...] because the different meanings shown in the guide words might share some common characteristics."
- (on run-in entries:) "Some words are not given in their alphabetical place because the different parts of speech of a word are grouped together when they share a similar form and meaning [...] they do not always have a separate definition."
- (on sense:) "When a single word has a meaning which differs slightly from the definition, this is shown or explained in an example." The *How to use the dictionary* pages also confirm that "Meanings that are slightly different from the main definition are explained [either] by a word or phrase in round brackets (=...) within an example sentence [or] by a complete sentence, not in italics, among the example sentences."

On the basis of the above we can establish that it is not just technical terms such as *polysemy* or *homonymy*, which might discourage the reader, have been omitted from the *How to...* sections, but that these

notions themselves were not considered when designing the entries. Just as you can't recognize homonymous words by Guide Words accompanying them, you can't hope to identify uniformly numbered senses linked via polysemy within entries, for the • sign often separates not senses, but one example sentence from another (as in the case of the noun **beam** LIGHT above), or end-of-entry quotations from the rest of the entry (as at the end of the same entry). Thus, within the same entry, **rage** ANGER gets explained as '(a period of) extreme or violent anger,' and exemplified not only by *Her sudden* ***towering*** *rages were hard to understand* • *I was frightened because I had never seen him* ***in*** *such* ***a*** *rage before* • and *She reacted with rage to his suggestion* • but also by '*(dated) This hairstyle* ***is (all) the rage*** (=very popular at the moment)'. The 'dated' sense has been smuggled in without a definition: it ought to be explained and/or the sense numbered. It is odd that CIDE, which usually spreads out the senses of polysemous words in different entries with a Guide Word, should not do this.

I find it highly regrettable that CIDE also uses the "oralized" definitions introduced by CCELD. These definitions sound like specimens of natural-sounding spoken English, and whatever their merit may be in the classroom and however useful they are for the teacher, they give rise to a host of problems. This is perhaps the one feature of the Collins Cobuild tradition which has come in for the heaviest criticism. What makes this especially precarious is that while in CCELD this is the only defining technique, here it is featured simultaneously with others, including the other type of "substitute" definition, the round-bracketed expressions. Thus, in the entry for **beam** LIGHT *v,* one sense of the verb is explained as "(=shone brightly)", another as "(=broadcast)". In the **beam** WOOD *n* entry, however, we read two sentential definitions, one nominal and one verbal[19] one: ... "**the** beam is a wooden bar..." and "... If you are on your **beam-ends**..." This means, then, that there are altogether four kinds of defining technique frequently in disguise in CIDE:

(i) the "standard" definition that always follows the main headword;
(ii) the Guide Word, which often acts as a synonym, but sometimes as a general term, actually repeated in the text of the standard definition;
(iii) the sentential explanations; and
(iv) the round-bracketed paraphrases.

It only makes matters worse that (although the adverb with the former veils this fact) both **shine** and **broadcast** are really *synonyms,* not

definitions, and even if these synonymous expressions may belong to the controlled defining vocabulary, they are certainly not simpler than the word defined. It is worrying to think of all those words in the dictionary which are thus not properly defined or whose definitions may be no more than synonyms.

This unfortunate closeness to CCELD in terms of the oralized, complete-sentence definitions is so perfect that CIDE even employs the word *also* as a discourse device (which, as such, is only legitimate as part of some *oral* explanation) to link to the definition of a new sense. In the entry **temper** BEHAVIOUR, the sense 'the usual state of your feelings which makes you become angry easily or stay calm' gets illustrated by five sentences separated by the • sign, followed by this explanation: 'Your temper is also the way you are feeling at a particular time,' illustrated by *He's* ***in*** ***a*** *bad/good temper...* and *Ask her when she's* ***in*** ***a*** *better temper.* This is followed by the definition 'Temper is also a tendency to become angry easily.' If you read these definitions and examples as self-contained fragments of the entry (which would deserve a number subentry in most other dictionaries), then the word *also* has obviously no right to be there.

If the editors have been willing to sacrifice so much — tradition, convention, rigour — to the immediate didactic needs, then one can only sympathize with them, and hope that the decision proves to be a right one.

4.1.2 The defining vocabulary

CIDE lets the reader in on facts about its controlled defining vocabulary of less than 2 000 words more than any other dictionary. Moreover, on page 1702 we learn that the *Language Portraits* and the *Usage Notes* are also written using these words, whose selection was based on the following principles:

- The words should be
  a. common, of high frequency
  b. easy for learners to understand
  c. useful for explaining other words
  d. those with the same meaning in British and American English;
- The words should not be
  e. confusable with foreign words
  f. among those often confused with other English words
  g. old-fashioned.

It might be difficult for the reader to see how (c) differs from (b), what it is that makes a word useful in explanations, or even what creates simplicity of understanding, but a short "user's manual" like this is hardly the place to clarify. Confusables, archaisms, potential false friends (for 16 languages, at least), and all kinds of troublesome words, to be sure, are guaranteed to be *out,* and the reader can check because the list itself[20] is provided on pp. 1702–1707. Problems show up, however, when the reader does check, and finds, for example, **history** listed, an item in whose entry in the body of the dictionary six country codes are listed warning of the nations where the word is a False Friend (see **4.3**).

4.1.3 Lexical relations

Synonyms and antonyms are not mentioned anywhere either in the promotion material or the *How to...* sections. Although in CIDE there is nothing comparable, *e.g.,* to the *Special Columns* of CCELD, or to LDCE's usage notes which also provide synonymous and antonymous phrases, they nevertheless appear in several disguises. Of these, it is presumably only the first two or perhaps also (*c*) and (*d*) below that are justifiable ways of providing synonym information:

a. in a few of the *Language Portraits,* to most of which the reader is guided from the entries; such are, for instance, the following, where this information is presented in the fashion of a thesaurus (these are all the examples from the M–O section of the *Pictures, Language Portraits, and lists of False Friends* page): *Measurements, Meeting someone, Memory, Money; One, Opposites.*

b. when an entry provides synonymous words *following* the definition: when, for example, **astonish** is defined thus: 'to surprise very much...; to AMAZE' Similarly, when at the end of entries • sign is followed by *Compare.*

c. when an entry offers a (near-)synonym[21] as *part of* its definition, *e.g.,* when **fury** is explained as 'extreme anger', or **rage** as '[...] extreme or violent anger'.

d. when the Guide Word is also a (near-)synonym[22] of the headword, for example: **beam** SMILE *v* 'to smile with obvious pleasure' (where 'smile' gets used twice within the entry, practically in the definition).

e. when a synonym is given *in place of* a definition, as in the case of **shine** and **broadcast** (see above).

4.2 Guide Words

In the previous section we have already touched on the function of the Guide Words. Here, I will be concerned with what types of words qualify as Guide Words in what types of entry. To be reminded of the function of the Guide Words, let us consider the five entries for **last**. They are as follows: (1) **last** [FINAL], (2) **last** [MOST RECENT], (3) **last** [CONTINUE], (4) **last** [NO MORE] and (5) **last** [UNSUITABLE]. It will be apparent that the Guide Words provide an immediate visual clue as to the difference between sense of the same word, between what are called the "core meanings."

The kinds of words that appear in the Guide Word boxes are sometimes synonyms (such as when **beam** [LIGHT] is distinguished from the word **beam** [SMILE]), often general concepts (in which—most fortunate—case they are like the classifiers[23] of definitions themselves, as when **lash** is marked by the Guide Word [HAIR]), while at other times just any word from the relevant lexical area (*e.g.,* when **aerial** [AIR] *n* gets contrasted with **aerial** [RADIO] *n*). It also happens that the word being looked up and the Guide Word belong to different parts of speech (as in the fourth example of **last** above), or when the commonest sense of **late** is distinguished from the rest by a preposition: **late** [AFTER].

An area where rigorous uniformity surely cannot be expected is the border regions between grammar and lexicon. Out of the 25 separate **do** entries, only 19 Guide Word boxes contain single verbs, another two are **do** [CAUSE TO HAVE] and the pseudo-verbal **do** [BE ACCEPTABLE] (where—if there is no restriction on what part of speech to use—I suspect the adjective would serve the purpose just as well); the rest are marked as follows: **do** [FOR QUESTIONS/NEGATIVES], **do** [FOR EMPHASIS], **do** [TO AVOID REPEATING] and **do** [PRISON], the only nominal item.

It is certain that consistency is lacking here; it is equally certain that the editors are well aware of this extravagance (uniformity cannot possibly be—neither is it—claimed anywhere). The other problem with the Guide Words is that some of them are simply not helpful, a few distracting, or even misleading. **Stand**, for example, has eight verbal entries, and while some of these only show the weaknesses discussed above (**stand** [VERTICAL], **stand** [PLACE], **stand** [POLITICS] and **stand** [HEIGHT]), others are not immediately recognizable even for the native speaker. I doubt that most speakers of English sufficiently familiar with the methodology of CIDE would be able to provide an example of **stand** [STATE] or **stand** [SUCCEED], and if that is so, then it is not something that the average learner can be expected to be guided by.

One cannot help thinking that most of the problems that the Guide Words (may) have solved have been caused by several other design features of the dictionary: by the destruction of the system of phrasal, prepositional, and phrasal–prepositional verbs, by the *Phrase Index*, and ironically, the Guide Words themselves.

Let me illustrate what I mean on the case of multi-word verbs. The word **up** has 23 separate main entries, some with indented subentries (LDCE has four, OALD just one). Clearly, something is amiss here, and it is not difficult to see what: (**a**) One of the Guide Word boxes with which **up**—an ungradable predicative-only adjective!—is entered is ROAD, meaning '(of a road) being repaired and so not suitable for use,' as *in The council has got the road up because of a broken sewer;* (**b**) another box—for the ungradable adverb **up**—has SMALLER, with the meaning given as 'broken or cut into smaller pieces; made smaller in area', illustrated by *The car blew up* and *Cut the magazines up...*; (**c**) a third has **up** TRIAL—this, too an ungradable predicative adjective!—with the meaning 'on trial in a court', illustrated as ... *he'll be up* ***before*** *the magistrate* • *Max is up* ***for*** *armed robbery;* while in (**d**) the alleged meaning of **up** AGE—another ungradable adverb—is 'to a greater age', exemplified by both • *No one said that* ***growing*** *up would be easy or painless* • *Many single parents struggle to* ***bring*** *their children up on a low income.*

The questions here are so serious that it seemed more convenient to mark the four entries with a letter and comment on them separately.

a. I do not think ROAD would be a good Guide Word even if the separate entering of this **up** were justified because I do not believe that many native speakers will associate the word *up* with roads except for *dig up;* I do not think that *up* is an adjective here, or in *dig up;* still less that *up*—or any other adverb (particle) can have the meaning 'being repaired and so not suitable for use.' Not only is this counter-intuitive: if this were what *up* meant, then **get** in the same verb complex **get up** or **tear up** would deserve a separate entry, with a Guide Word, say ROAD, of its own, and a meaning along the lines of 'dig up (a road) for repair to make it suitable for use.' That, I think, sounds absurd enough.[24] By the way, the supposed predicative adjective **up** should not occur in **got the road up** exactly because it is marked as [after *v*], *i.e.*, 'adjective that only follows a linking verb.'

b. I do not think the two examples here—**blow up** and **cut up**—have to do with one another: they may share the element 'completely'

(this would be the sense of this *up* in most dictionaries), but the first has a spatial aspect that the second one does not; SMALLER is a rather clumsy way of capturing *both* the result of explosion and an a cutting spree; if *up* in *cut up* means 'cut into smaller pieces,' then the reader wonders what the meaning of *cut* might be.

c. I do not think that the *up* part of either *be up before (a court/ magistrate)* or *be up for* [*some crime*] has a sense of and on its own: out of context, *Max was up* can never mean that he was in a court, magistrates' or other. Any context would not even do: my informants tell me that dialogues like ***A.*** *They found out about the robbery* ***B.*** *Is Max up?* are not possible, so it is only the entire multi-word verb itself that can qualify as context here. I also do not think that *up* is an adjective here, and I fail to see what the criteria are that make adjectival an *ungradable predicative-only adjective* that is always used with *be,* and is not gradable.

d. **Up** in ***growing*** *up would be easy* and ***bring*** *their children up* shows very much the same problems (which I will not repeat), but it does not mark the adverb as ungradable. It may be a good idea to mark each adverb particle as a *nongradable* adverb: what it means, however, for an adverb to be nongradable, never gets explained. The *Grammar labels* list has the label [not gradable] referring to the LP⟩ *Comparing and grading,* but this panel says not a single word about *adverbs,* just adjectives.

The remaining question, then, is whether the user will profit from all this simplification or will be discouraged from resorting to the Guide Words if s/he thinks them haphazard and undependable. I myself would like to believe that the over-simplification does not overshadow usability, but even so I am convinced that a second edition, for example, without a thorough revision of all these aspects is impossible.

5 Above the word

5.1 Selection and collocation

CIDE provides selectional information not as a list of collocates for words, in some coded or skeletal form. The type of subjects and objects collocating with a given verb, or the type of noun collocating with an adjective, is specified by an *of*-phrase in many dictionaries; this is not done here either. No uniform system of indicating these resembling ODCIE 1983 is used, where, *e.g.,* in the entry **the beam in one's own eye**, the following information is given between definition and examples:

V: (not) see, notice; ignore; remove.

Generally, CIDE provides most of selectional and collocational information in and by the examples, here too guided evidently by the suspicion that whatever comes the reader's way in the form of a full phrase or sentence will be remembered better than codes, letters and lists of isolated words.

If one uses full phrases and sentences rather than isolated words in the fashion of the entry **the beam in one's own eye** because one is unwilling to tamper with one's corpus, then often a separate sentence has to be included just for the sake of one minor grammatical or lexical variation. If no editing of the corpus sentences is insisted on, then in the **beam** case this would mean finding in the corpus and entering one example each for the lexical varieties **see the beam**..., **not see the beam**..., **notice the beam**..., **ignore the beam**... and **remove the beam** IN ONE'S OWN EYE(S). After all, no corpus sentence can possibly contain **see** and **not see the beam** or any other verb with and without negation at the same time. (Though it obviously draws from a huge corpus, ODCIE claims no "authenticity" in terms of unedited examples, *cf.* **3.4** above). If, however, you opt for whole sentences for *didactic reasons,* then you can include all the wealth of the corpus because at least the device of using slashes between alternative items is legitimate. In the above case this will yield something like **Irene tries to see/to notice/not to ignore the beam in her own eye(s)**. It is obvious that considerations of space will soften up the principles of even those who otherwise prefer 100% unedited examples. It is equally easy to see that this example is much more difficult to decipher than the self-contained sentences (each with a possibly long—but truly authentic—name in subject position).

The presentation of variation in grammatical pattern is fraught with similar difficulties (*cf.* Fillmore 1989). CIDE appears to be guided by the usual principles of clarity and simplicity when, for example, it enters the verb **blame** in an entry with one meaning only,[25] and informs about its syntagmatic combinability simply by means of the example sentences separated by the • sign:

> ... *don't blame me... if you miss the bus* • ... *I blame the parents* • ... *blames his mother* **for** *his lack of confidence* • ... *blames his lack of confidence* **on** *his mother* • ... *campers were blamed for start***ing** *the forest fires* • *the hot weather is partly* **to blame for**... *the water shortage* • ... *'a bad workman blames his tools'*

It is noteworthy that—in all entries—the examples themselves are not set in dark type so that not only idioms but also the most important grammatical features can be highlighted in this way.

I also wanted to check **load** for its dual pattern **load sg on(to) sg** (as in the famous *load hay onto the wagon*) and **load sg with sg** (*e.g., load the wagon with hay*) but I was disappointed to find that in CIDE *load* only means 'put into', and you can only load X *into* or *onto* Y, moreover, that X must be something operated by a piece of equipment, and Y, this piece of equipment. The patterns **load a gun** and **load the film** are given without their respective prepositional phrases as well, but here, the former obviously does not fit the definition — the gun itself is not put anywhere. The "mixed" pattern **load the gun with bullets**, on the other hand, is not featured.

As regards the notorious problem of the alphabetical placement of collocations of all kinds, CIDE, which has devised the *Phrase Index,* seems to have taken two bold steps which will hopefully prove popular with users. Noun–verb (*i.e.,* subject–verb), verb–noun (*i.e.,* verb–object), and adjective–noun combinations can theoretically be entered under two headwords, whether of the non-idiomatic or idiomatic type. For production purposes,[26] all of them would have to be entered under the noun, while for the purpose of comprehension such as reading in a foreign language, each — adjective, noun and verb alike — belongs in its alphabetical place. Now, number one: CIDE makes no distinction between idiom, collocation, phrase, combination and the rest anywhere in the *How to...* notes, abandoning a separation that the average user has never been fully aware of and may always have felt uncomfortable about. (The *Usage Labels* list does not contain terms like "idiom" or "idiomatic", either). Number two: CIDE gathers them indiscriminately in the *Phrase Index,* which contains such diverse items as **BBC pronunciation** and **hold/keep at bay**, **beach resort** and **Belisha beacon**, **beam at** and **beam with** as well as **bear down** and **beat one's breast/chest**, to enable the user to check here first and always find the word, fast.

5.2 Idioms: The *Phrase Index*

The *How to find words and meanings* pages state that although "[some phrases and combinations of words] are usually explained in examples following the definition for the *first* word of the combination (for example 'dead end' is found at **dead**, and not at **end**)," the *Phrase Index* is useful

if you want to check where a combination is explained. The short guide at the top of the *Phrase Index* itself has this:

> over my dead **body** 350R33
> over my **dead** body 350R33
> **over** my dead body 350R33

indicating that you find this expression looking for any of the most important words, moreover, a line reference is also provided. (This is particularly useful with long entries; 350R33, *e.g.,* directs the reader to line 33 in the right column of page 350). The *Phrase Index,* then, gathers together word groups under the letter of each of their components. The function of boldface here is to show which of these components is the word which justifies entering the given phrase under the particular letter (so following the items down a column means a bit of a zigzag reading task). For example, the expression **grin and bear** it is entered twice, both under the letter B and under G, as follows (about stress marking see section **6** below):

> In B: ▫grin and ▪**bear** it 624L66
> In G: ▫**grin** and ▪bear it 624L66

The function of the *Phrase Index* is to facilitate idiom lookup, "giving instant access to 30 000 phrases and idioms". The term "phrase" is to be understood very loosely, including all kinds of grammatical and lexical collocations as well as phrasal verbs in the broad sense; together with idioms, this Index ought to list all multi-word combinations — idiomatic and non-idiomatic — that feature in the entries — see the short list in the last paragraph of **5.1** above. However well represented they may be within the entries (and in those entries that I have checked, it appears that care has been taken to enter many more such items under the noun headword than in similar dictionaries), subject–verb and verb–object combinations are not among the "phrases" in the *Phrase Index.* It is mostly adjective–noun collocations of many types, noun–modifying-noun units, phrasal, prepositional, and phrasal–prepositional verbs, complex prepositional expressions such as **beat about/around the bush**, compounds like **world-beater**, and idioms such **fight a losing battle** or **fight like cats and dogs** that this Index contains. In its present form, *Word combinations* or *Words in groups* would perhaps be a better title.

There are several problems with the *Phrase Index.* One of the snags is that some of the items here need no inclusion at all. If you find a hyphenated compound (like **world-beater**), you are not likely to look for it under the letter of the second element. Multi-word verbs, are a more

complicated case, but they too, can generally be entered under the verb, not under the particle, even less under the preposition. (*Be* may be a special exception in several dictionaries because of the poverty of its own meaning, but all the others never come under the particle/preposition; indeed, CIDE has no multi-word verb with *be* entered under **be** itself).

It is hard to guess what percentage is meant by the word "usually" in the wording at the first paragraph of this section. Given, however, the almost limitless capabilities of the computer for arrangement and rearrangement, one wonders whether it would be a better strategy to enter combinations in the entry for their *first* word *always*—not just "usually". Two remarks must be made here. First, lookup time is significantly reduced only in the case of more-than-two-word expressions; with those consisting of just two, it does not really matter whether you open CIDE on a trial-and-error basis and go straight to one of the components, or start at the *Phrase Index*. In the former case, you stand a fifty per cent theoretical chance of being right, which saves you a second check; in the latter case, you always wind up checking twice. The other problem is that the entries themselves in the body of CIDE have now been pruned of the idiom cross-references: neither **body** nor **over** contains a boldfaced **over my dead body**.

This is an especially acute question if we observe that the *Index* (pp. 1708–1771) uses up more than three per cent of the 1770-page dictionary, which would produce an extra fifty pages of body text. (It may, however, be argued that the resultant removal of the cross-references from the entries has freed up at least that much).

A practical consideration against it, in its present form, is that the small print of the five-column(!) *Index* makes it—particularly the number-letter-number codes like the one above—hardly legible. Some people may find **boldface** especially difficult to distinguish. The use of the (primary and secondary) stress marking *here* would be seriously called into question even if it were not for legibility problems.

5.3 A word on/in "real English"

The latest advances in pragmatics and discourse analysis were noticeable in both the 1987 new edition of LDCE and CCELD published in the same year (Carter 1989 : 38). Authentic language in authentic situations is an ideal primarily for communicative language teaching, and in all dictionaries there will be entries that lend themselves more easily to this approach than others. The not-quite-standard "filler" *like,* the tactfully disagreeing *actually,* the discoursal connector *now,* the interjection *well* (together with

practically all of what are termed interjections, including swearwords), as well as "expressions" like *you see* or *the thing is* are cases in point. But if their discourse-orientedness makes them a happy hunting ground for the communicative classroom (these words were *just not in* before this method appeared on the scene), there should be available in or near that classroom a dictionary which describes and exemplifies them as fully as possible, for these words require a lot of contextual support.[27] I have chosen the word **shit** to illustrate this point because CIDE offers an unparalleled quantity of collocations and idioms in this entry as well as in similar ones — "real English" at its best is doubtless being shown here — and because a brief look at this entry in other dictionaries also promised to be instructive, and showed the superiority of CIDE in this domain.[28] I am not suggesting, of course, that the inclusion of all types of slang, taboo or nonstandard items makes a dictionary more communicatively-based or guarantees it to be useful in the teaching process. The stress on *natural, colloquial* use and *spoken* English — partly made possible by the huge database — works towards that end. If we follow the development of this entry, we can observe how dictionaries have been adding more encoding features by offering not just word and definition but phrasal and sentential examples and collocations/idioms, as well as by distinguishing between "descriptive" and "discoursal" uses.

CCELD 1987 has five collocations: **tough** ∼, **in the** ∼, **be** ∼**ting oneself**, *doesn't* **give a** ∼ *about,* and **beat/kick/knock the** ∼ **out of**. There is only one example sentence, which means that coded information like EXCLAM or N COUNT: ALSO VOC in the *Special Columns* is all that is given about the *use* of the word. This dictionary hardly registers the fact that the word *shit* is very much part of very informal spoken English. The 1978 LDCE has just two collocations, while the 1987 edition provides another two, giving one example sentence in the noun entry, and entering the *interj taboo* **shit** separately ("expressing anger or annoyance"). While the second edition of OALD has just one sentential example, the fourth edition already has four, including the *interj (taboo sl)* — altogether eleven examples and expressions. This is the background against which CIDE's *n taboo slang* stands out, with the following 17 collocations are listed (just with the noun!), almost all of them in one example sentence. The *exclamation taboo* word is separately entered and also exemplified in two sentences.

is a little ∼; is a load of ∼; what's all this ∼ about...? get a lot of ∼ from; doesn't take (any) ∼ from; doesn't know ∼ about; doesn't give a ∼ about; have/get one's ∼ together; no ∼! have ∼ for brains; when (the) ∼ hits the fan/flies; beat/knock/kick the ∼ out of; scare/frighten/ terrify the ∼ out of; with a ∼-eating grin; be/get on sy's ∼ list; be a ∼-stirrer; ∼-stirring; See also: BULL∼, SHITE

6 A note on transcription

This section is simply a list of the problems to do with transcription that I think need to be addressed. Some concern British *vs.* American pronunciation; one, the marking of stress in the *Phrase Index;* the third just alerts to a typographical error.

A word about the symbols *£* and *$* in the transcription: while it is appreciable that they are shorter than *Br* or *BrE* and *Am* or *AmE* would be, uniformity is impaired by their use because in other regional labels, *Br* and *Am* are actually used, along with *Aus, Irish Eng, Scot Eng* and *Canadian Eng*(!).

I have found that the $ and especially the £ sign standing for British and American in the transcriptions are printed too close to the actual symbols, which may be disturbing particularly when the word begins with a vowel: the British **amorous** is transcribed as /£ˈæm·ə·rəs/ and **o'clock** is /£əˈklɒk/. Simply italicizing would probably help.

The American pronunciation of **girl** is /$ gɜrl/ in the *How to use...* section, while in the body of the dictionary it is /$ gɜːrl/. Although in CIDE, the former must be a misprint and the latter one is the standard in all such words—/$ fɑːrm/, /$ hɔːrs/—it is the short vowel that actually reflects the phonetic reality better.[29]

Even if it were not for legibility problems, I would seriously doubt that the *Phrase Index* is the ideal place for (primary and secondary) stress marking of the items included therein. The use of the empty and black boxes, furthermore, added to the very small print and the difficult job of deciphering what the bold face stands for (*cf.* **5.2**), is most disturbing. The following is a specimen which I hope truly shows the size as well:

■**beach** □buggy 107L22 and □**beach** ■resort 107L26

The *Varieties of English • Differences in pronunciation* panel on p. 1609 contains a regrettable typographical error. The transcription symbol for the BrE sound **eə** = AmE **e** is illustrated on the three words **hair**, **fare** and **wear**, all of which are transcribed using the symbol **ea** instead of **eə**: hea^{r}, fea^{r} and wea^{r}. The wrong symbol — this time in larger bold type — also appears in *Phonetic symbols* at the back, in the word **hear**, again transcribed as /£ hea^{r}/.

7 Summary

No learner's dictionary can afford to be *just a dictionary* in the traditional sense (in which it may still be thought of by many people): a list of words with explanations. Ideally, it unites the good qualities of a textbook, an encyclopedia, a thesaurus, an atlas, a guide to errors and pronunciation, a slang dictionary and a specialized dictionary (of a field usually depending on the interest of the user), a basic grammar course and possibly other things, in one volume, which is preferably thin and light and inexpensive but up-to-date, comprehensive and like a CD disk as regards speed of information access. I think that in many of these important senses CIDE meets very high expectations, thanks to all of its design features, the *Guide Words,* its extra material, (to a smaller extent) the *Phrase Index,* and the numbered lines.[30]

I also believe, however, that no dictionary can afford to be *just a learner's dictionary.* For many people, that category has an amateurish feel about it, suggesting a thin booklet with lots of imprecision, babytalk, coloured illustrations coupled, alas, with childish lexicographic content. Editors and buyers alike seek to strike a balance between technical precision and authenticity, academic accuracy and intelligibility, exacting formal analysis of language and ease of use. We have seen several signs of CIDE being aware of the importance of this balance, from the *False Friends* through *Language Portraits* and *Pictures* to the checking for factual accuracy of entries in special subject areas, which they claim is a practice so far uncommon in EFL dictionaries. The encyclopedic character has been apparently added without jeopardizing intelligibility. I feel that most of the criticisms above have been targeted at things which would not exist if practicality had not been the highest priority.

CIDE is advertised as an international dictionary for students of English, a production as well as a decoding dictionary, explicitly recommended as a teaching resource: some of the *Language Portraits* "will provide enough material for a whole lesson." I do think that it does not fall

short of expectations in these respects, either. The least that can be said about it is that it is impossible to ignore the existence of CIDE: English lexicography is not what is was before it.

NOTES

[1] Landau 1984 is still one of the most authoritative sources with a comprehensive account of dictionary types.

[2] The "user/learner" category only makes sense with a dictionary for learners, in our case, an ESL/EFL dictionary. There can be little doubt that CIDE is such a dictionary, so we can probably say that most of its users are learners — this, however, by no means guarantees that it is a dictionary which they actually use in order to learn English.

[3] It would be hard to evaluate CIDE, and probably any other dictionary, without occasional reference to the competitors' publications. This is exactly what we will do.

[4] At the time of the very first edition of the OALD in 1948, the term valency from dependency grammar, which these Verb Patterns elaborate, was not a dirty word yet. This is where the notions complementation or subcategorization goes back to, and as linguistic theory is getting more and more "lexicalist", they have a more and more respectable ring about them.

[5] Now, for the first time, it has also entered the newest edition of the OALD.

[6] "Phrase" is exceptionally elusive. It is not even clear whether the word refers to the same thing in this as in the previous paragraph, *cf.* **5.1** and **5.2.**

[7] Incidentally, the *Oxford Advanced Learner's Dictionary*, now in its "new fifth edition", is also advertised as one that uses "the results of key language research and extensive consultation with teachers", and using which "millions of students around the world have improved their written and spoken English".

[8] CCELD has 1703 pages, ends in Zulu, and has not a single page of back matter; the OALD dictionaries have always included a great amount of such material at the back; LDCE seems to be taking a middle position.

[9] And there may exist questionnaires that I have no knowledge of, indicating that this indeed is what the majority of users prefer.

[10] It is odd that *cursor* should not yet be included in CCELD.

[11] A word is only given a reference if "the place where it is explained is more than five dictionary entries away from where you expect to find it" [*i.e.,* alphabetically] (CIDE p. ix).

[12] See the previous Note. This usually affects run-ins but not cases like *firefighter* ~ *fireman.*

[13] The 5th edition of OALD is now being advertised as "The dictionary that really teaches English."

[14] In association with the University of Cambridge Local Examinations Syndicate.

[15] A similar — and similarly imprudent — strategy is to have the "Uncountable" label [U] follow rather than precede the relevant meaning.

[16] Placing these indented subentries within the same entry is also know as nesting or clustering.

[17] It seems that not only affixed derivatives are run on in indented subentries, but all items which differ in word class membership from the headword and "share a similar meaning." There is, of course, a sense in which this too is derivation — zero affixation — but having unnumbered senses run on is by no means conventional.

[18] We are assuming, for practical purposes, that homophony and polysemy do exist, even though we are aware that they have fuzzy boundaries, at two ends of a scale. It is perhaps interesting to note the paradox that CIDE, clearly for reasons of simplifying, has gone further scrapping the distinction. As for *beam,* this word is a particularly illustrative example of the uncertainty of these terms. CCELD, but not CIDE, contains an example which shows this: *he beamed his thanks* is clearly on the boundary between 'broadcast' and 'smile'.

[19] The nominal variety is perhaps less disturbing, because the initial "In the sport of women's gymnastics" can be read as if it were a register label, and the copula being a light element the whole string does not have to really read like a sentence.

[20] This list is on six pages, in a six-column layout, approx. 120 lines per column, *i.e.,* more than 4 300 items. This includes derivatives (*e.g., add, addition, additional, additionally*), words with different senses (*address* 'home details' and *address* 'speak to').

[21] In many cases, of course, it may hard to decide whether the definition contains a near-synonym (*anger*) specified by another word (*extreme*), or a hyperonym of the headword; according to the definition, after all, rage is a kind of anger.

[22] It is again possible to maintain either that (near-)synonymy holds, or that this is a case of hyponymy with *beam* defined through the genus *smile* plus the differentia; by the definition, *beam v* is a kind of *smile v.*

[23] Classifiers as opposed to distinguishers: the general term — genus — as opposed to the differentia.

[24] Other dictionaries, too, use explanations like these with *up,* but hidden deep within the entry this is much less disturbing than with the small separate entries each with a Guide Word.

[25] LDCE has two meanings: 'consider (sy or sg) responsible for (sg bad)' and 'find fault with'.

[26] All—not just learner's—dictionaries are used both for production/encoding and comprehension/decoding purposes, but in EFL dictionaries the former element dominates.

[27] Especially when we note that unlike in the classroom, where intonation, paralinguistic features, even gestures are there to help, the dictionary has to compensate for all of that by means of the examples.

[28] The definition, incidentally, shows several deficiencies: the labels [U] and [C] following, not preceding the noun in question (discussed in **5.4**); the mixed ("oralized" and "standard") character of the definitions (discussed in **4.1**); and the more serious confusion between "mention" and "use" of one and the same item, as in (disapproving) "Shit can also be used to refer to someone or something you do not like. . . ," or (disapproving) "Shit is also insults, criticism or unkind or unfair treatment," where the first words of the example sentences should be highlighted (set in upper case in CIDE).

[29] Most editions of OALD do not provide separate American pronunciations at all, and the 1987 edition of CCELD does not either. LDCE, on the other hand, also prints both the /r/ and the broad vowel in the American version of these words. Interestingly, J. C. Wells's *Pronunciation Dictionary* also marks this type of vowel as 'long'.

[30] One indeed wonders whether the numbered lines—and many of the possibilities offered by the coputer—may not be put to better use in some way.

REFERENCES

Bakos Ferenc, ed. (1994) *Idegen szavak és kifejezések kéziszótára* [*A Dictionary of Foreign Words and Phrases*]. Budapest: Akadémiai Kiadó.

Bookseller 1995 = The Bookseller, 3 February 1995

Carter, Ronald (1989) 'Review of Longman Dictionary of Contemporary English & Collins Cobuild English Language Dictionary'. *International Journal of Lexicography* **2:1.** 30–43.

CCELD = John Sinclair, editor in chief (1988) *Collins Cobuild English Language Dictionary.*

CIDE 1995 = Paul Procter editor-in-chief (1995) *Cambridge International Dictionary of English.* Cambridge: CUP.

CUPPR 1995 = 'Oldest Publishers Launch Streetwise Dictionary: Lighhearted Press Release For Tabloid Media And Chat Shows'. 16 February 1995. Cambridge: CUP.

Fillmore, Charles J. (1989) 'Two dictionaries'. *International Journal of Lexicography* **2:1.** 57–83.

Hausmann, Franz Josef and Adeline Gorbahn (1989) 'Cobuild and LDOCE II: A comparative review'. In *International Journal of Lexicography* **2:1.** 44–56.

Landau, Sidney I. (1984) *Dictionaries: The Art and Craft of Lexicography.* New York: The Scribner Press.

LDCE 1978 = Paul Procter, ed. (1978) *Longman Dictionary of Contemporary English (1st ed.).* Harlow: Longman.

LDCE 1987 = Della Summers, editorial director (1990) *Longman Dictionary of Contemporary English (New edition, 11th impression).* Harlow: Longman.

LDELC 1992 = Della Summers, editorial director (1992) *Dictionary of English Language and Culture.* Harlow: Longman.

NFC 1995 = *New From Cambridge—CIDE promotion brochure.*

OALD 95 = Jonathan Crowther, editor (1995) *Oxford Advanced Learner's Dictionary (5th ed.).* OUP.

ODCIE 1983 = A. P. Cowie, R. Mackin and I. R. McCaig (1983) *Oxford Dictionary of Current Idiomatic English. Vol. 2: Phrase, Clause & Sentence Idioms.* ???: OUP.

Procter, Paul (1995) 'A new dictionary you can read in the bath'. Information leaflet. Cambridge: CUP.