

Quote index unquote

David Crystal

This article discusses the indexing principles and procedures employed in compiling Words on words: quotations about language and languages, by David and Hilary Crystal, published by Penguin in May 2000.

This year sees the publication of one of the most enjoyable books it has been my good fortune to write – or rather help to compile, for it is a book of quotations; a book of quotations about language, to be exact. It is a co-compilation, in which my own reading has been complemented by that of my wife, who also played a major role in the preparation of the index. Nobody can read everything; and to have two intuitions available for such a work immeasurably improves its coverage.

To say it is a compilation does not diminish the creativity involved in the exercise. The actual amount of creative writing is indeed limited, being restricted to the thematic organization, the introduction, the editorial interpolations which contextualize a quotation, and of course the index. But with books of quotations the creativity goes well beyond the writing; the very task of selecting a quotation is a creative one. For, in the vast majority of cases, it is the compiler who decides whether a series of words is to count as a quotation or not. That is why it does not sound inappropriate to say that someone is the ‘author’ of a book of quotations, rather than the ‘editor’.

The point is sometimes forgotten or unnoticed. Writers do not decide whether their sentences become quotations. Even if they write their books as sets of potential quotations – as in books of maxims – it is not up to them. Only readers decide. In some cases, a consensus among readers quickly emerges, and we end up with a universally acknowledged quotation of the ‘To be or not to be’ type. In other cases, a quotation may exist only in the form of a marginal mark or textual underline in one’s own copy. People compiling books of quotations aim to record all of the first kind; but if they restrict themselves to these, their books will not be very large. The vast majority of any compilation consists of quotations that are not universally known. Indeed, one of the pleasures of reading a book of quotations lies precisely in discovering those observations that one never knew before. The discovery is the thing. It is common for a compiler to ‘contract out’ the work – to ask someone to read through an author’s oeuvre, and to find 10 or 20 putative quotations. For those who know an author very well, this task is by no means easy. If you love Wordsworth, the whole of his work is quotable. Being restricted to even 100 quotations presents the Wordsworthian enthusiast with a remarkably difficult task.

Compilers therefore need to be clear in their minds about their selection criteria. This not only helps the reader understand why some items are included and others are not; it is a prerequisite for deciding what to include in the index. The first issue, of course, is that of coverage. What is the topic of

the book of quotations to be? If the answer is ‘everything’ – as in the major collections of general quotations – then there is no real problem. Virtually any word in the selected item is a possible source of an index entry. But if the compilation aims to deal with a restricted domain of knowledge, as in the present case, then the situation is different. A quotation about language may – and often does – contain a great deal of extraneous subject matter, which is part of the context of the quotation but dispensable, in the sense that the linguistic point being made exists without dependence upon it. This material does not need to be indexed. An example is the opening two lines of this famous quote about words from T. S. Eliot’s *Four quartets*; they have to be there, in order to grasp what is being said, but they are nothing to do with the topic of language as such:

So here I am, in the middle way, having had twenty years –
Twenty years largely wasted, the years of l’entre deux guerres –
Trying to learn to use words, and every attempt
Is a wholly new start, and a different kind of failure
Because one has only learnt to get the better of words
For the thing one no longer has to say, or the way in which
One is no longer disposed to say it.

In the index of key words, phrases and concepts (see below), accordingly, only those items are included that are specifically related to language within the quotations (or their associated context). Incidental comments, observations and descriptions that happen to be part of the quotations are not indexed. In this way, the index avoids becoming a concordance of all the words in the book.

Coverage

What is contained within the language domain? For the present article, this is most easily summarized by listing the headings of the 65 thematic sections into which the book is organized (Table 1). This shows that *Words on words* is very much about language in a linguistic and literary sense, including several areas of application (such as translation and language teaching) and incorporating a number of topics from the field of communication in general (such as body language and the media). The book tries to deal in a balanced way with language in general as well as languages in particular, covering each of the four main channels of language (speaking, listening, reading and writing), and also the various branches of language structure represented alongside the main contexts of language in use.

Why language? The work is not intended solely for the professional linguist or literary specialist. Language is one of

Table 1. The thematic headings, and their glosses, recognized in *Words on words***LANGUAGE**

- 1 **The nature of language** The nature and functions of language
- 2 **Language in thinking and thought** Thinking and reasoning in relation to speech and language
- 3 **Language myths and origins** The origins and early history of speech and language
- 4 **Body language** Facial expressions, gestures, and other forms of non-verbal communication
- 5 **Meaning and sense** The nature of meaning, definition, sense, and nonsense
- 6 **Words or deeds** The choice between talk and action, language and living
- 7 **Ever-changing language** Continuity and change in language over time; loan words and neologisms
- 8 **Arguing about usage** Issues of usage and correctness in language
- 9 **The language of youth and age** Child language, and the changes between youth and old age

LANGUAGES

- 10 **Language diversity** The world's languages; language identity, endangerment, maintenance, and death
- 11 **Languages observed** The supposed qualities of different languages
- 12 **Bilingualism and multilingualism** Being proficient in more than one language
- 13 **Translating and interpreting** The task of translation and interpreting
- 14 **Teaching and learning** Principles and practices of language teaching and learning
- 15 **Foreigner talk** Foreign impressions and usages

ANALYSING LANGUAGE

- 16 **Exposing language** Terms and techniques for analysing language; linguistics and philology
- 17 **Speaking** The nature and functions of speech and speaking
- 18 **Pronunciation** Speech sounds, vocal organs, and the sound of the voice
- 19 **Listening** The nature and functions of listening and hearing
- 20 **Reading** The nature and functions of reading and books
- 21 **Writing** The nature and functions of writing; techniques of successful writing
- 22 **Learning to read and write** Literacy and orthography, spelling and its reform, writing systems
- 23 **Grammar** Grammar, grammars, and points of grammatical usage
- 24 **Dictionaries** Dictionaries and lexicographers

GOOD AND BAD LANGUAGE

- 25 **The nature of eloquence** Oratory, eloquence, rhetoric, and other forms of good or bad expression
- 26 **The art of conversation** Conversational topics and strategies
- 27 **Subject-matter** Topics, knowledge, opinions, and beliefs as expressed by language
- 28 **Language clear and unclear** Obscurity, vagueness, imprecision, ambiguity, and other issues of clarity
- 29 **Saying too much** Verbosity, long-windedness, and other forms of excessive speech or writing
- 30 **Saying just enough** Brevity, taciturnity, conciseness, and other forms of economy in speech or writing
- 31 **Keeping quiet** The nature and functions of silence and pause
- 32 **Friendly language** Words of comfort, love, apology, gentleness, praise, and tact
- 33 **The language of flattery** Flattering, smooth, and insincere language

- 34 **Unfriendly language** Speaking or writing sharply; sarcasm, ridicule, anger, insult, and slander
- 35 **Quarrels and arguments** Quarrelling, disputing, arguing, and debating
- 36 **Truth and lies** Telling the truth, lying, secrets, vows, and other promises

WORDS

- 37 **Words, words, words** Words in general, their nature and function
- 38 **Words praised** Words and language praised or celebrated
- 39 **Words criticized** Words and language criticized or condemned
- 40 **Words as weapons** Words seen as weapons or ammunition
- 41 **Wise and foolish talk** Words of wisdom or advice; words careless or foolish
- 42 **Slang** Slang and fashionable language
- 43 **Swearing** Swearing, cursing, obscenity, and other exclamatory language
- 44 **Worn-out words** Clichés, platitudes, euphemisms, and other worn-out expressions
- 45 **Names and nicknames** Names and nicknames of people, places, or animals
- 46 **Words and expressions** Individual words and expressions praised or condemned

STYLE, GENRE, AND VARIETY

- 47 **The secret of style** The nature of style in language
- 48 **Language in literature** The use of language in literature and criticism
- 49 **Poetic language** The distinctive language of poetry; the differences between poetry and prose
- 50 **Metaphors and similes** The use of images in speech and writing
- 51 **Accents and dialects** Regional ways of talking, in pronunciation, grammar, vocabulary, and style
- 52 **Noblesse oblige** Language in society; social class and politeness
- 53 **Men and women** Opinions and attitudes about male and female speech
- 54 **Walls of words** Academic language, and jargon in general
- 55 **The language of politics** The jargon of politics, diplomacy, the civil service, and public administration
- 56 **The language of law** Legal terms, personalities, and procedures
- 57 **The language of religion** Talking to or about God, and other forms of religious language
- 58 **The language of science** The language used in science and technology
- 59 **The language of the media** Communication in the press, advertising, radio, or television
- 60 **The performing arts** The language of theatre, cinema, or mime
- 61 **The musical world** Language in relation to music, opera and song
- 62 **Proverbial wisdom** Proverbs, maxims, epigrams, and other succinct expressions
- 63 **Playing with language** Puns, jokes, anagrams, wit, and other kinds of language play
- 64 **Personalities** Linguistically distinctive characters in fact and fiction

POSTSCRIPT

- 65 **Quoting and misquoting** The use and abuse of quotations

those topics with enormous popular appeal – for it is the stock-in-trade of orators, media personalities and writers of all kinds. An apt quotation about pronunciation, slang, or style is likely to enhance many a lay opinion, whether the context be a school essay, a radio quiz show, a political speech, a novel, or a letter to the press complaining about a point of usage. When we encounter language quotations from such different worlds as Plato and Winnie the Pooh, Shakespeare and James Bond, it is evident that we are dealing here with a domain whose interest and relevance is universal – probably far greater than the miscellaneous topics that are already well served by quotation collections, and that go by such intriguing titles as *Read my lips* (for politics) and *Rave on* (for popular music).

What counts as a language quotation? We used several criteria.

- In view of the general nature of the readership of this book, the subject matter had to be of a correspondingly general kind – an observation about the nature of language as a whole, or about one of its widely recognized branches or topics, such as speaking, reading, usage, eloquence, slang and grammar.
- We were looking for memorability, in the sense that an extract had already achieved some currency within the English-language-using community – a kind of linguistic equivalent of ‘To be or not to be’. (There were very few of these.)
- We looked for relevant items that had actually been used as quotations, in the press, in textbooks, and elsewhere (though often misquoted).
- We looked for succinctness and autonomy of expression. An ideal quotation, for us, would be a single sentence, with a compelling rhythmical structure, making a point which stands alone without need of any further context. There were actually very few of these, too, other than proverbial expressions – as illustrated by the fact that we had to make regular use of square brackets at the beginning of a quotation, giving the context for what follows.
- We looked for observations that relied on vivid or unusual metaphors, similes and other figurative expressions. There were plenty of these, especially in the literary domain. On occasion, we found fine imaginative treatments of a topic which extended over several sentences, and we included these as wholes – the boundary between a quotation and an anecdote is often difficult to draw.
- We tried to avoid banal observations, of the ‘Language is complicated’ type.
- We found we had a penchant for the ironic, satirical, humorous or playful observations about language, and there turned out to be quite a large number of these.

Table 2 gives a selection of instances, illustrating these criteria (apart from the penultimate one).

Where have the quotations in this book come from? From many different sources. We do not believe in rediscovering the wheel. We therefore began by reading all the standard anthologies of quotations, to see what they had already picked up on the language topic. We then chose certain key

Table 2. Examples of quotations illustrating the six positive criteria of selection

General observation

Language seeks vengeance on those who cripple it.
(George Steiner, 1967, ‘The retreat from the word’, in *Language and silence*)

Memorability

Language is fossil poetry.
(Ralph Waldo Emerson, 1844, ‘The poet’, in *Essays: second series*)

Frequently quoted

Talking jaw-jaw is always better than war-war.
(Winston Churchill, 26 June 1954, speech at The White House, reported in *New York Times* the following day)

Succinctness

Slang is a language that rolls up its sleeves, spits on its hands and goes to work.
(Carl Sandburg, 13 February 1959, *New York Times*)

Figurative

Words form the thread on which we string our experiences.
(Aldous Huxley, 1937, *The olive tree*)

Playful

A sentence may be defined as a group of words, uttered in sequence, but without logical connection, to express an opinion or an emotion. A number of sentences if emitted without interruption becomes a conversation. A conversation prolonged over an hour or more becomes a gossip. A gossip, when shared by several persons, is known as a secret. A secret is anything known by a large and constantly increasing number of persons.
(Christopher Morley, 1919, ‘Syntax for cynics’, in *Mince pie*)

authors and worked systematically through them – if ‘work’ is the right word for having an excuse to read or re-read all of Shakespeare, Dickens, Wordsworth, Twain and others. Several authors are known for their observations on language, such as Johnson, Emerson, Hazlitt, Butler, Wilde, Castiglione and Montaigne, and their essays received special attention. One source led to another, because observations about language, at a given time, were often made in response to others. The 19th-century arguments about the origins of language are a case in point, bringing together a cluster of famous names, such as Darwin, Müller and Whitney, as well as some famous extracts (such as the Psammetichus origins-of-language experiment) which, because of their standing in the history of linguistic thought, demanded more space than the average quotation. Some sources were read with high expectations but yielded next to nothing – such as Pepys’s *Diary*. Others were unexpectedly fruitful, such as the works of Laurence Sterne.

We would of course like to have read everything, but we were working within a space constraint of 150,000 words (an underestimate, as it turned out) and a time frame of just over a year, from conception to publisher’s deadline. We were under no illusions, then, that this book would in many ways be only a first bite at the language quotations apple. In particular, we were conscious of the way we inevitably focused on English-language authors. Parallel books could be compiled using the literature of any language, and we hope they will appear. In the meantime, although we included (translated) items by authors of some 40 nationali-

ties, it is obvious that there is a vast amount more in the leading authors of other cultures than our own.

Organization

Books of quotations use different organizational criteria. Most list their quotations by author, in alphabetical order. This is the standard approach in such works as Oxford (Partington, 1996) and Chambers (Jones, 1996). But a thematic or semantic approach – such as Hoyt (1909), Tripp (1987) and Cohen (1998) – has just as respectable a history, and for a restricted domain, such as language, where the topics are more the focus of attention than the authors, this seemed to be the more illuminating approach. Note the assumption: we feel that people who approach this book with a query in mind will be more interested in the topic than its author. ‘I need to find a good quotation on slang for a speech or article’ or ‘I’m trying to remember who made that observation about slang’. We believe these to be more likely motivations than ‘What did Wordsworth say about language?’ Of course, we have to allow for the latter enquiry, too – which is what motivates an Index of Authors, listing every quotation from an individual writer throughout the book (see Table 3, and below). But that, we believe, is a secondary motivation.

Opting for a thematic organization, however, brings its own difficulties. In particular, one has to decide what to do when a quotation could be appropriately assigned to more than one theme. For example, Sir Arthur Helps’s witty remark (in *Friends in council*, 1847, II, Ch. 1), ‘Reading is sometimes an ingenious device for avoiding thought’, could equally well be placed in the section on ‘Language in thinking and thought’ or on ‘Reading’. Some editors (e.g. Cohen, 1998) adopt a repetition solution when faced with such decisions and simply repeat the quotation under the different thematic headings. We did not like this solution; it takes up space that could be more valuably used for different quotations, and we feel it short-changes the reader. But if we do not repeat, then we must make a decision about single-theme assignment, and must cross-refer. This is what we did. So, we placed the above quotation at

Table 3. The first few items in the Index of Authors, illustrating the style used

Abercrombie, David (1909–92) British phonetician	51:1
Achard, Marcel (1899–1974) French playwright	53:1
Achebe, Chinua (1930–) Nigerian novelist	44:1, 62:1
Acheson, Dean (1893–1971) American statesman and lawyer	21:1, 64:1
Adamov, Arthur (1908–70) Russian-born French playwright	37:1, 44:2
Adams, Douglas (1952–) British novelist	13:1–2, 16:1, 23:1, 30:1, 45:1, 46:1
Adams, Henry (1838–1918) American historian	2:1, 51:2, 55:1, 64:2
Adams, John (1735–1826) American statesman and second president	10:1
Adams, Robert M (1915–) American academic and literary critic	16:2, 30:2
Addison, Joseph (1672–1719), English essayist and politician	1:1, 2:2, 11:1–2, 15:1, 18:1, 21:2, 27:1–2, 31:1, 38:1, 39:1, 49:1–2, 50:1, 57:1, 59:1–2, 61:1–2, 63:1–3

‘Reading’, cross-referring at the end of the ‘Thought’ section. Every thematic section therefore has a list of cross-references to the items in other sections that contain relevant material.

Words on words contains one other kind of cross-reference. Occasionally, we would notice a close similarity between two quotations – often because one author was consciously using or adapting the words of another. For example, Coleridge says, in his *Biographia literaria* (1817, Ch. 17): ‘Anterior to cultivation, the lingua communis of every country, as Dante has well observed, exists every where in parts, and no where as a whole’. This obviously requires a cross-reference to the relevant Dante quotation, which occurs in *De vulgari eloquentia* (I, Ch. 9). These cross-references are placed immediately after the relevant items.

Within themes, we decided to follow conventional practice by ordering individual quotations alphabetically by author, and then (if more than one quotation appears from a single author) by date. Sequences of proverbs, however, required a different approach (there being no author), and these we ordered alphabetically on the basis of what seemed to be the most relevant keyword (e.g. a proverb about *speech* appears before one about *words*). This semantic approach seemed preferable to one that ordered proverbs on some other basis, such as geographical or linguistic origin, as it enabled us to group all proverbs to do with a particular topic together, thereby allowing cultural parallels and differences to be seen. This method is used, for example, in Selwyn Champion’s major collection, *Racial proverbs* (Champion, 1938).

After each individual quotation we gave its source, beginning with the author’s name (where this is relevant), the date of the quotation (where known), and the location. Within each quotation, we tried to provide continuous extracts, using ellipsis dots only to avoid unnecessary discourse-connecting words such as *thus* or *however*, or to eliminate an authorial parenthesis or amplification. Quite often we found it necessary to give some relevant context for a quotation, placed before it within square brackets. This was especially useful for quotations from plays or novels, where we needed to identify the speaker (and often the addressee). The topic to which a quotation relates also sometimes had to be explained, and as far as possible we did this using a paraphrase of the author’s own words.

Indexing

The book has three indexes: an Index of Authors; an Index of Sources; and an Index of Key Words, Phrases, and Concepts. The Index of Authors has already been illustrated: it is arranged letter-by-letter and gives dates, nationalities and professions. The dates are to enable readers to see the period within which an author was writing: for example, observations about English as an emerging world language vary in their impact, depending on the century in which they are made. The nationalities and professions add an important contextual dimension; *what* is said is sometimes not as important as *who* said it. Essentially the same observation can take on very different implications if it comes from a poet, a professional linguist, or a film star – or,

Table 4. The first few items from the Index of Sources, illustrating the style used

<i>ABC of Reading, The</i> , Ezra Pound, 5:50, 27:36
<i>Absalom and Achitophel</i> , John Dryden, 29:28, 49:37
<i>Absentee, The</i> , Maria Edgeworth, 6:16
'Absolute and Abitofhell', Ronald Knox, 57:30
<i>Act of Creation, The</i> , Arthur Koestler, 2:32, 2:67–70, 39:28
<i>Acts of the Apostles</i> , 12:3
<i>Adam Bede</i> , George Eliot, 26:34
<i>Ad Pammachium</i> , St Jerome, 30:29

to continue the above example, whether a claim about world English is being made by a speaker of English or French.

We were not sure just how much use would be made of an Index of Sources (see Table 4), but felt that in some cases it could provide a motivation for enquiry. The books of the Bible are one case in point; the plays of Shakespeare are another. We can also readily conceive of someone wanting to find out the relevant language references in Boswell's *Life of Johnson* or Wordsworth's *Prelude*, and doubtless there are other such cases. As always, indexers are trying to predict the future. We set the titles of books, plays, large poetic works, independent essays, periodicals and newspapers in italics; titles of individual poems, journal articles and contributory essays in inverted commas; and untitled items (such as letters to the press) in roman without inverted commas. The order of this index is also letter-by-letter.

But the bulk of the indexing work related to the need to provide coverage of all key words, phrases and concepts. Because most of the quotations are not immediately recognizable, in the way that 'To be or not to be' is, an index of first lines, or famous metaphors, or key words in context, is not the way to handle the material in this book. Of course, if a quotation is well known, to a recognized group of likely readers, then it must be handled accordingly. Thus, 'all grammars leak' is indexed as a whole (as well as under *grammars* and *leak*), because this remark is famous among linguists – a typical examination question (followed by 'Discuss'). But most of the items are not like this. As a result, the general index looks much more like a conventional book index, with important concepts selected and grouped, so that people who want to find quotations relevant to a particular topic (other than those evident from the sectional grouping into general themes) will be able to do so. Some quite interesting groups of subheadings emerge in this way. The approach also proves to be the most useful when trying to track down a quotation that has been translated into English, and where the actual words used in the translation can only represent (rather than uniquely identify) the words found in the original. The alphabetical organization for this index, suiting its semantic purpose, is therefore word-by-word.

The general index, accordingly, turned out to be much larger than the indexes usually found in books of quotations – almost twice as large as those found in other works within this genre.¹ It was designed with two different aims in mind. First, and most obviously, it is there to help people find a quotation that they already know (but are unclear about its source or author) or – far more important – that they *think* they know. If our own experience is typical, people regularly half-remember quotations. The problem for the indexer is

that there is no way of predicting which half of the quotation will be recalled. All the key words and phrases in a quotation therefore have to be seen as potential access points, and the index must include them all. For example, to handle the Emerson quotation, 'Language is a city to the building of which every human being brought a stone' (from his 1875 essay, 'Quotation and originality'), the index must anticipate access points under *city*, *building*, *humans* and *stone*, as well as *language*. Only in this way can the index help the user who says 'What was that quotation about language being a *N*?' – where *N* could be any one of these words. If a quotation is several lines long, as many are, it can be seen that the number of potential access points is very large.

Secondly, the general index has to help people find a set of quotations which they do not already know but which they are looking for in order to meet a particular need (such as preparing a speech, article or lecture). Here, the thematic divisions of the book are an important resource, but the 65 divisions we have recognized do not go far enough to meet the needs of people who want to know whether there are any language quotations about, say, Hebrew, heaven, or hyphens, or whether words have ever been compared to bees, birds, or boats. Or, if you are a dentist, and you want to know if there are any quotations about language and dentists (there are – two). Or, if you want to find something to do with tennis (there are seven) or horses (there are nine). An in-depth alphabetical index is the most convenient way of aiding such searches. An illustration of the depth to which we have gone is given in Table 5.

The notion of 'key words' is reasonably transparent, and we have applied it as literally as possible, replicating in the index the words that appeared in the quotation, both as headwords (which one would expect) and also in the sub-entries. The words are as close to what appears in the original as we can get, within the syntactic constraints of indexing procedure. Sometimes, we repeat the relevant section of the quotation, word for word, as in these entries under 'slang':

as a linguistic luxury 42:18
 as a poor-man's poetry 42:19
 as blank cheques of intellectual bankruptcy 44:11

Sometimes, there is a selection and simplification, identifying just the lexeme (or word type) crucial to the identity of the quotation. In the case of Milton's 'The troublesome and modern bondage of rhyming' (from the preface to *Paradise lost*), our judgement is that the crucial lexemes are *rhyme* (which underlies *rhyming*) and *bondage*, not the adjectives *troublesome* and *modern*; hence our reciprocal entries as:

bondage, . . .
 of rhyming 49:76
 rhyme, . . .
 bondage of 49:76

On the other hand, the adjective is crucial in the well-known quotations about puns as 'battered witticisms' (Oliver Wendell Holmes), 'false wit' (Addison) or 'the lowest kind of wit' (Dryden). Hence the fuller phrasing in the first few lines of the entries at 'puns':

admitted among the smaller excellences 63:8
 analogy with lightning 63:21

as a pistol let off at the ear 63:22
 as battered witticisms 63:20
 as false wit 63:1
 as good or bad things 63:14
 [=clenches] as the lowest kind of wit 63:16

In all cases, the aim is to give enough information for the quotation to be recognizable, and no more.

But the general index could not be restricted to identifying only the most salient words. This is one of the main limitations of the indexing approaches used in most books of quotations, which focus on individual words without taking into account their collocations. Often, the identity of the quotation depends on the whole of a phrase, rather than on an individual word, as we have just seen. But, as an extension of this principle, if the identity of a quotation depends on a whole clause or sentence, this needs to be indexed too. That is why 'All grammars leak' (Edward Sapir) is indexed under *all* (as well as under *grammar* and *leak*) and 'Let thy words be few' (Bible) is indexed under *let* as well as *words* and *few*. It is very difficult to extend this principle beyond the single-clause sentence: if the quotable item is too lengthy – as in a multi-clause sentence, or a sequence of sentences (as in some of Shakespeare's dialogues on the subject of words) – the index would quickly get out of control.

There are many places where the important thing is to index the concept or underlying issue involved, as well as the key words. For example, under 'pun', we have an entry 'Johnson's aversion to'. A conceptual approach to key words

Table 5. The first few entries from the *d*-section of the entries under *words* in the key words index (references are to numbered items within thematic sections)

dallying with 39:55
 dancing on a plume of 64:19
 dazzling the eyes 38:13
 dead when said 17:9
 dead without sense 5:34
 dearest 1:16
 dealing destruction through time 40:11
 death of 7:17, 37:44
 debased and hard 7:10
 debates on 8:53
 decaying 28:21
 deceiving us 37:40
 decent to exclaim in 43:49
 deeds not 6:16
 defaced by long usage 39:29
 delicious to the nose and throat 37:39
 delight of 20:38
 derivations of 8:66, 22:25
 desirable permanency of 27:25
 destroying instincts 2:142
 destruction of (beauty in) 37:47
 details covered up (in Latin words) 28:43
 dignity of 7:17
 diminish things 39:30
 diminution of the lungs by corrosion of 30:55
 disappearing 7:15, 38:37
 disbenching 40:48
 discovering 14:32
 discussion of 2:5
 disparaging 34:13
 dissecting 37:12
 distrust of 39:28
 divested of their magic 38:39

also allows the use of paraphrase, to enable the salient point to be the focus of the entry. A frequent example here is the use of comparison or analogy: 'words' are compared to all kinds of things, in this literature – bricks, sweets, gems, leaves, and so on. The exact way in which the comparison is made – whether as metaphor or simile, or the form of syntax used – is not the issue. The important thing is to note the nature of the analogy. Hence index entries are often a paraphrase of the original quotation: 'words, analogy with arrows'; 'words, as instruments'.

Recognizing both key words and concepts, of course, means that homonyms must be distinguished. It would not be sensible, in a semantically aware index, to list all instances of, say, *tongue*, under the one heading. *Tongue* has two relevant senses, as a vocal organ and as a language. Quotations relating to the different senses are therefore rigorously distinguished. The same applies to *readers* ('books' vs 'people'), *English* ('language' vs 'people'), *mean* ('poor' vs 'sense'), and many more. Semantic distinctions of this kind are not normally introduced into quotation indexes. In the Oxford compilation, for example, the entry 'English as she is spoke' is adjacent in the list of sub-entries to 'English have hot-water bottles', and 'If ever I do a mean action' is next to 'It all depends what you mean'. When lists of sub-entries are short, this might not be much of a problem; but when the lists are long, conflating major senses should definitely be avoided.

An interesting problem arises with quotation indexes: what to do with archaic and dialect forms – spellings, punctuation, word endings and syntax? Take this quotation from Chaucer's *Troilus and Criseyde* (V, 1786): 'And for ther is so gret diversite / In Englissh and in wrytyng of oure tonge'. Do we index *diversite* as it is or do we modernize it to *diversity*, and list it along with all the other references to this topic? Our semantic approach argues for the latter: it is a historical accident, in a sense, that *diversite* has changed its spelling. The important thing is the nature of the allusion that Chaucer is making. That is what we want the index to capture. We do have to recognize, though, that sometimes an archaic or dialect identity is critical to a quotation as a whole. In such cases it is important to retain the original spelling throughout. The linguistic identity of quotations from Robert Burns, for example, needs to be respected: *plain braid lallans* (= 'broad Lowland Scots') is therefore listed under *braid* and *lallans*.

Another technical problem is what to do with words, word elements and phrases used as citations from an individual language, when these are homonymous with general topics. For example, there is an entry relating to *Rosalind* as a character who speaks the words of one quotation and another relating to *Rosalind* as a name when it is the focus of a quotation. A second example is *fancy* as a topic, which needs to be distinguished from *fancy* as a preferred noun for the topic. This is a standard problem in linguistics, and in books on this subject a convention has long been present of using italics for any instances of actual language used, and roman for terms that talk about the language (the metalanguage). This is our practice, too. Actual language expressions are always placed after their homonymous metalinguistic equivalents.

The only item we decided not to index as a headword, on grounds of practicality, was the subject of the book,

'language(s)' – though we do include it in a handful of cases where it is part of an established quotation (as in Johnson's 'Language is the dress of thought'), and it is of course included as a sub-entry within entries. Any quotations using the word *language* are therefore indexed using its collocations. Were there any other words which we might have excluded in this way? Having completed the indexing of 'word(s)', which came out at 1428 lines, it may be debated whether it was a sensible decision not to exclude that item also. But, having done the work, we decided to leave it in.

The general index was an interesting exercise, because of its experimental nature. It is a synthesis of three types of information that are not usually distinguished in books of quotations: whole quotations; key words and phrases that are part of real quotations; and relevant concepts underlying the whole or part of a quotation. As a result, the index grew rapidly in length, and we were worried that the publisher would have a problem with it. In the event, thanks to an economical design and a genuine interest on the part of Martin Toseland, our Penguin editor, to 'do the job properly', as he put it, we were able to present the index without needing to cut it.

The result, nonetheless, was much larger than we had anticipated at the outset. In its final form, *Words on words* contained just under 3900 quotations, which took up 298 pages. There are just under 30,000 entries in the three indexes, taking up 280 pages. It has been a new experience for me, compiling an index that is virtually half of a book. But to deal with quotations properly, there is really no other way.

Note

- 1 Comparisons are difficult, because of variations in typography and layout; but if one uses a within-book ratio, then 26% of Chambers's pages are devoted to its general index, as are 28% of Oxford's pages. In our book, 48% of the pages are devoted to the general index.

References

- Champion, S. G. (ed.) (1938) *Racial proverbs: a selection of the world's proverbs arranged linguistically*. London: Routledge.
- Cohen, M. J. (ed.) (1998) *The Penguin thesaurus of quotations*. Harmondsworth: Penguin.
- Hoyt, J. K. (ed.) (1909) *The cyclopedia of practical quotations*. London and New York: Funk and Wagnalls.
- Jones, Alison (ed.) (1996) *Chambers dictionary of quotations*. Edinburgh: Chambers.
- Partington, Angela (ed.) (1996) *The Oxford dictionary of quotations*, revised 4th edn. Oxford: Oxford University Press.
- Tripp, Rhoda Thomas (ed.) (1987) *The international thesaurus of quotations*. New York: Harper and Row.

David Crystal is a writer, lecturer and broadcaster on language and linguistics, a reference books editor, and an honorary professor of linguistics at the University of Wales, Bangor. Editor of The Cambridge Encyclopedia 'family', his recent authorial writing relates to the theme of the world's dying languages in the book *Language death* (Cambridge University Press, 2000). He is a former honorary president of the Society of Indexers. Address: Akaroa, Gors Avenue, Holyhead LL65 1PB, UK; email: crystal@dia1.pipex.com