

Key elements of dictionaries and other language references

THE ENTRY TERM

We turn now to a discussion of the elements that make up a dictionary entry, beginning with the alphabetized headword, or main entry, by which the word or expression being defined (the lexical unit) is identified. The canonical form, sometimes called the *lemma*, is the form chosen to represent a paradigm; most headwords, with the exception of cross-references and names, are canonical forms. (For a discussion of the selection of entries in dictionaries, see Ch. 7, p. 356.) This section will deal with the question of determining the form and placement of the headword for each lexical unit.

In order to have canonical forms, forms that the speakers of a language recognize as representative of grammatical paradigms, there must be a standard language. If there are competing forms with exactly the same meaning, one must arrive at some basis for deciding which of the various usages is to be represented in the dictionary as the canonical form. Thus, before a dictionary can be written for a language, the language must have developed more or less standard spellings or, in a language with various dialects, have a preferred dialect. Variant spellings and dialectal forms can, of course, be given, and for the larger (and especially the historical) dictionaries should be given; but a single form must be chosen as the canonical one.

In English, the standard, which emerged during the fifteenth century, was that of the East Midland district that included London. Spelling remained variable throughout the sixteenth century, one of the chief aims of the early English lexicographers, such as William Bullokar and Richard Mulcaster, being to bring about greater uniformity in spelling. Though they directed attention to the problem, they did not succeed in resolving it. In the early eighteenth century individual variations in spelling were widely denounced as intolerable, but it was not until Bailey's dictionaries of the 1730s and more particularly Johnson's of

1755 that the spellings of many words became fixed. In America, Noah Webster was an active and influential spelling reformer, as already noted.¹

"The first task of the editor of a dictionary," writes James R. Hulbert, "is to decide on the spelling of his word-entry. Usually on a modern dictionary this affords no difficulty, as usage has fixed a single spelling."² Spelling, however, is not the only problem. In bilingual dictionaries, for example,

One of the traditions of Western lexicography is to use the so-called "infinitive" form for both the entry heading and the translation of verbs . . . In many parts of the world, verbs are usually entered under the plain present (or non-past) form, and it is misleading to translate such headings with English *to* + constructions. Japanese *suru* does not mean "to do"; it means "(someone) does" or "will do." The one advantage of the *to* + translation is that it clearly marks the word as a verb, and in English many verb forms are homonymous with nouns.³

Because the users of bilingual and ESL dictionaries may not know the canonical forms used as headwords, it is particularly important in these dictionaries that inflected forms that differ markedly from the canonical forms, such as *is* and *went*, be listed as headwords in their own alphabetical positions with cross-references to the canonical forms (*be* and *go* in the examples cited). If space permits, even more closely related inflections, such as *made* (*make*) and *tried* (*try*) should also be listed.

The traditional practice of large, comprehensive dictionaries, whether historical like the *OED* or synchronic like the American unabridged dictionaries of the nineteenth to mid-twentieth century, was to list virtually all canonical forms as main entries. Even if a word had the same form in more than one part-of-speech, each part-of-speech appeared as a separate headword. College-level and shorter dictionaries, pressed to conserve space, often combine all parts-of-speech within a single dictionary entry, the notable exception being the Merriam-Webster college dictionaries, which do not.

Some linguists call the overall organization of the dictionary, as determined by the headwords (or lemmas), the *macrostructure* of a dictionary, and the organization of the information within each dictionary article the *microstructure*.⁴ The particular way in which definitions and other features of the dictionary article are presented comprise the microstructure. Are definitions arranged by part-of-speech? Are definition numbers used? Are subsenses distinguished by alphabetic letters also included? Are examples included? If so, in what form? How are cross-references indicated?

Homonymy and polysemy

One aspect of the decision as to whether or not to list a lexical unit as a headword has to do with whether it is perceived as belonging to the same cluster of meanings as other lexical units of the same form, or whether it has a completely different meaning. If the form and grammatical function of two lexical units are the same but the meanings are different, are we dealing simply with two different meanings of the same word (*polysemy*) or with two different words? *Homonymy* is the condition of two words that are pronounced alike or have the same spelling, irrespective of grammatical function, but that differ significantly in other respects. *Wind* (a current of air) and *wind* (to turn something around) are homonyms, as are *lie* (to say something untrue) and *lie* (to place oneself in a reclining position). Homonyms are further distinguished by whether they are spelled the same (*homographs*) or pronounced the same (*homophones*). Thus, both pairs of words cited above are homographs, but only the second pair (*lie*) are homophones. In alphabetically arranged dictionaries, spelling rules when determining headword status, so as a practical matter lexicographers are concerned with only with those homonyms spelled alike, not with homophones that are spelled differently, like *way* and *weigh*. They are concerned about words like *light*, which as a noun means a source of illumination and as an adjective means weighing comparatively little. How does one determine homonymy and distinguish it from polysemy?

Determining homonymy is important because most dictionaries accord homonyms separate headword status, whereas definitions in a polysemous entry are usually listed together, although, as we shall see below, this is not always the case. Homonyms become part of the macrostructure of the dictionary, whereas polysemy usually affects the microstructure. Though long discussed and debated by linguists and lexicographers, there is no consensus about how to determine homonymy.⁵ Although etymology is often mentioned as a criterion, it is an uncertain guide, as etymologically disparate words have sometimes evolved associated meanings, and words that can be traced back to the same root have often evolved distinct meanings, such that the modern speaker does not regard them as being related.⁶ Zgusta sensibly remarks, "Homonymy begins at the point when the speakers of a language are unable to conceive different senses as connected . . ."⁷ Divergence of sense must be the overriding factor in determining homonymy, and there is no sure standard by which to determine in every case whether the

divergence is sufficient to warrant separate headword status as a homonym rather than status as a definition in a polysemous entry. In bilingual dictionaries, there are other considerations as well: "With languages having overt gender markings, . . . practical arguments favour separate entries for homonyms of different genders."⁸

In the day-to-day practice of lexicography, especially commercial lexicography, no one is concerned about the theoretical imperatives of distinguishing between homonymy and polysemy. The editors are solely concerned with representing the language in a way that will be seen as sensible and comprehensible by the dictionary user. Accordingly, some types of dictionaries, notably school dictionaries and ESL dictionaries, often do not distinguish the meanings of homonyms from other meanings that are polysemous senses of the headword. This is distressing to some scholars who have studied homonymy. One comments, "Thereby the differences between homonymy and polysemy are effaced, which misrepresents the real situation in the language."⁹ But there is no "real situation" in the language. The situation is whatever the lexicographers believe it to be, based on their best understanding of both the language and the needs of their intended audience. To make matters more complicated yet, some ESL dictionaries, such as the *Cambridge International Dictionary of English* (CIDE) and the *Cambridge Dictionary of American English* (CDAE), accord separate headword status to different meaning clusters. This treatment obscures rather than emphasizes homonymic status, but it was seen as a way to help foreign learners find particular meanings, since learners lack the knowledge and intuitive sense of native speakers that make them "unable to conceive different senses as connected."

Run-on entries

In order to conserve space, virtually all dictionaries, even very large ones, have adopted the practice of "running on" at the end of entries the canonical forms of grammatically related words. Many regularly formed adverbs and adjectives, called *run-on entries* (or *run-on derivatives*), are thus listed without definitions at the end of the entry for an adjective, noun, or verb. Thus *racially* may be run on to *racial* and *transferable* to *transfer*. The presumption is that if one knows the meaning of the entry word and of the suffix – which is defined in its own alphabetic place as a main entry – one will have no difficulty understanding the derivative word.

In an effort to cram as many entries as possible into a dictionary –

since the run-on derivatives also count as entries – many dictionaries include thousands of rarely used derivatives, such as *oppressingly*, *sluggardliness*, and *idioticalness*, which appear in the *Collins English Dictionary (CED)*. Some run-on derivatives may never have been used. Words that could exist but for which no record exists to prove that they have ever been used are called *latent words*.¹⁰

Though the practice of using rare or latent words to enhance one's entry count is an unfortunate waste of space, it does not contribute to confusion or misunderstanding. But the failure to list semantically important words as main entries is more serious. *CED*, for example, runs on *oppression* to *oppress*. Given the heavy usage in recent years of *oppression* in various senses, this is an abuse of the presumption that derivatives do not require full definition.

The assumption that certain forms are more basic semantically because they happen to be more basic grammatically is not necessarily true. To regard adverbs ending in *-ly* or nouns in *-ness* as less important than the adjectival root indulges the convenience of the definer at the expense of the needs of the user. In many cases, *-ly* words and *-ness* words have acquired senses not adequately covered by the root words, and in some cases the *-ly* adverbs are far more commonly used than the adjectives to which they are run on. *-ly* is supposed to mean "in a – manner." *-ness* is supposed to mean "the state of being –." Thus *swiftly* and *swiftness*, run on to *swift*, can be understood to mean "in a swift manner" and "the state of being swift," respectively. Well and good. It is understood in theory, if not always observed in practice, that if a derivative has a meaning not covered by the senses of the form to which it is appended, or not applicable to the formulaic definitions cited, it should be entered separately and defined. As a result, adverbs like *hopefully*, *incidentally*, and *literally* are defined as main entries in all reputable dictionaries.

But the problem is much broader than that of defining derivatives separately because they include senses peculiar to themselves. To say that *oddly* means "in an odd manner" or that *obviously* – which is far more common than *obvious* – means "in an obvious manner" is both misleading and ambiguous. Yet this is routinely done to save space. A check of the Cambridge International Corpus shows that *obviously* occurs more than three times as often as *obvious* when speech is considered as well as written texts. (See Ch. 4, p. 192 and Ch. 6, p. 296 for information about corpus use in lexicography.) In written text, the frequency of *obviously* in both British and American English is only slightly less than that of *obvious*. In spoken material, *obviously*, which is often used as a sentence

modifier ("Obviously, I was wrong") occurs about four times as often as *obvious* in American English and more than ten times as often in British English. Yet in all the American college dictionaries except for *Merriam-Webster's Collegiate Dictionary*, Tenth Edition (*MW10*), as well as in *CED*, *obviously* is run on to *obvious*, without a definition. Similarly, *fortunately*, which is also often used as a sentence modifier and which in overall frequency is about as common as *fortunate*, is run on without definition in every one of the above dictionaries except *MW10*. *Completely*, though slightly less common in British and American writing than *complete*, is far more common in speech in both varieties, so that its overall frequency is much greater. Yet *completely* is run on to *complete* without a definition in all the American college dictionaries, including *MW10*, as well as in *CED*. The gloss "in a complete way" cannot be said to give an adequate explanation of *completely* in such common examples as "I was completely exhausted" or "I forgot my umbrella and got completely soaked." The meaning here is rather "to a very great extent" or "altogether." Lexicographic practice has in the past not allowed one to run on an adjective to an adverb, so the poor adverb, no matter how widely used, is pushed to the rear of the entry where it is lucky to get a pronunciation, much less a definition. However, some ESL dictionaries, such as *CIDE* and *CDAE*, do allow an adjective to be run on to an adverb. Nonetheless, adverbs remain a neglected part of speech. Traditional dictionaries seem to be bound by a kind of linguistic primogeniture, rewarding only the oldest and leanest of words with full treatment, while regarding derivative forms as tainted with the marks of an indelible subordination.

Webster's Third New International Dictionary (*NID3*), the last of the unabridged dictionaries, defines *oddly* as "in an odd manner or to an odd extent" and cites three illustrative quotations, while the entry for *odd* occupies nearly an entire column of type and is divided into six main senses subdivided into thirty-two subsenses. Are we to understand that *oddly* can mean "in an odd manner or to an odd extent" in each of the thirty-two senses of *odd*? Obviously not. More, in which senses of *odd* is *oddly* most commonly used, and in which is it rarely used? This information is central to an understanding of the use of *oddly*, but it is not given. Many *-ly* words fall into this category: *openly*, *modestly*, *deliberately*, *centrally*, *strangely*, and so on. For example, to say that *strangely*, in "He was strangely silent," means "in a strange manner" does not explain it. In this context it means "inexplicably" and is not related to *strange* in the sense of "unfamiliar or alien," as in "It was a strange environment for

her." Yet *NID3* defines *strangely* as "in a strange manner," without even an illustrative quotation, and all the American college dictionaries and *CED* run it on without definition. What makes these omissions peculiarly important is that the undefined words are very commonly used. I have suggested elsewhere that modern dictionaries have a bias in favor of scientific and technical terms, and I believe the neglect of common words such as adverbs is the other side of the coin.¹¹ Dictionaries have increasingly taken on an encyclopedic function that precludes full treatment of common words. Common adverbs are regarded as less important than even those scientific terms rarely used in general contexts.

In the past, the comparative frequency of particular words was unknown, so frequency could not be used as a consideration arguing for headword status. But with the development of large linguistic corpora, that is no longer the case. (See Ch. 6.) Relative frequency of particular senses in both speech and writing can now be discovered through an examination of very large collections of text stored electronically. Even very large citation files, because they are collected by a process of selection, cannot be used reliably for statistical studies of frequency because they are apt to be unrepresentative of the language as a whole. As James A. H. Murray had occasion to remark in connection with the *OED* files, citation readers all too often ignore common usages and give disproportionate attention to uncommon ones, as the seasoned birder thrills at a glimpse in the distance of a rare bird while the grass about him teems with ordinary domestic varieties that escape his notice. By contrast, a corpus that is sensibly developed will, by design, be representative, at least to a much greater degree than any citation file. There is no longer any excuse for dictionaries to ignore frequency when determining headword status. It is a far more reliable guide than a lexicographer's hunch.

I do not maintain that relatively high frequency ought to guarantee a word headword status, but it is obviously of great relevance and is one of the most important factors to be considered. I am not even sure that we ought always to be wedded to the canonical form as the headword if its inflections are far commoner, especially in dictionaries for foreign learners. Surely it is confusing for many learners to have to find the entry *be* for many examples including *is*, *are*, *was*, *were*, and *am*. Why not put the headword under a form the user is most likely to look up, such as *is*? Of course, a cross-reference at *be* would refer the readers who sought the word there to look under *is*, where one would find all the other inflected forms and the canonical form. Such exceptions would be extremely rare, limited to a few of the most common verbs in the language.

Dictionaries are already in a quandary about what to do with common comparatives of adjectives that have developed their own meanings, such as *worse* and *better*. These are inflections but are also considered as lemmas, and are usually given headword status, but with one definition referring to the word's status as a comparative or superlative. Thus, *worse* is identified as the comparative of *bad* and *ill* and then redundantly defined as "bad, evil, etc., to a greater degree," along with other definitions specifying particular ways of being bad to a greater degree. *Worse*, *more*, *most*, and *better*, *best* are just too common with too rich an assortment of meanings, connotations, and collocational environments to be treated simply as inflections. This is an entirely reasonable decision. Though the status of such adjectives is not exactly comparable to common verbal inflections like *is* – which no one would argue is a canonical form – the question it raises is the same: should the representation of the language's formal structure take precedence over the reader's ability to use the dictionary to best advantage? The answer depends on the type of dictionary and its intended audience.

Scientific nomenclature

Scientific and technical dictionaries have terminological difficulties of a different sort. A group of medical experts convened in 1968 to discuss the problems of medical terminology concluded:

Most diseases are described under several names, some under as many as 20 or 30 different synonyms. The same drug is often designated under numerous different names . . . Equal confusion reigns in such diverse fields as bacteriology, virology, chemistry, to name just a few of the basic sciences, and nearly all of the medical sciences. A Russian and a French psychiatrist, for example, could never be sure that when using the same term they were designating the same entity, and the same would be true of a German and an American dermatologist.¹²

The problem here is neither of spelling nor of paradigmatic model, but the absence of a standard nomenclature. It is not analogous to the lack of a standard language, as in a country with mutually unintelligible dialects. German and American dermatologists believe they understand perfectly well one another's words, but they attach different meanings to the same medical term. Two people speaking widely different dialects, on the other hand, know very well that they cannot understand each other. The differences in pronunciation, stress, and other features make the same word unintelligible, even though it has the same meaning in both dialects.

How does one go about standardizing nomenclature? In other words, how does one determine preferred terms when there is no agreement among the experts? Recommended criteria, in order of importance, are as follows:¹³

1. Usage, as in textbooks, medical dictionaries, and other reference works
2. Recommendations of authoritative specialist organizations
3. Self-descriptiveness, i.e., giving preference to terms that describe the nature of the concept, as opposed to eponymous terms (terms named after people, such as *Bright's disease*). However, some eponymous terms are so well established by usage that they should be retained.
4. Specificity, i.e., not vague or ambiguous
5. Simplicity, i.e., as short as possible
6. Conceptual relationships, i.e., related concepts should ideally be related in terminology. Two forms of pneumonia should both include the word *pneumonia*. This may seem self-evident to those unfamiliar with medical terminology, but it is very often not the case, sometimes for good historical reasons. For example, often the causes of a newly identified disease are unknown or misattributed. Thus, like many other disorders, *legionnaire's disease* was named after the group of people (*American Legion* members) who, it was mistakenly thought, were the first to contract it. One of the early names for AIDS (before *acquired immune deficiency syndrome* became the accepted name) was *gay immunocompromise syndrome*, because, when the condition was first described, it was associated only with gay men. A disease originally thought to be a form of tuberculosis and accordingly named may turn out to be more closely associated with pneumonia, but by that time the original name given to the disorder may be widely known and have been translated into numerous other languages. If the name of the disease were to be changed suddenly it would cause widespread confusion among scientists who had known it only by the original name – and had known very well that it was a form of pneumonia – and even more so among scientists of other countries, because the new translation might be quite dissimilar to the original one. In fact, it is impossible to obliterate forms in widespread usage, and the attempt to do so merely aggravates the problem by multiplying the number of competing usages in currency.
7. Linguistic relationships, i.e., ideally the translations of a term into other major languages should be similar in all languages, so that one can more easily recognize that the two terms are the same.

ALPHABETIZATION

Dictionaries usually alphabetize letter by letter rather than word by word. They place *power*, *powerful*, and *power of attorney* in that order, whereas a word-by-word arrangement would place *power of attorney* before *powerful*. Letter-by-letter arrangement has the great virtue that readers need not know whether a compound is spelled as one word, as a hyphenated word, or as two words. Since usage is often divided about compounds – witness *database* and *data base*, *e-mail* and *email* – and is constantly shifting, the ability to locate such terms is of considerable practical importance.

Verbal idioms are particularly troublesome. No form of alphabetization can successfully deal with all types of idioms without listing each in several places, and no dictionary can afford the luxury of such repetition. However, with the aid of a computerized database it is now possible to include an idioms index in which every idiom is listed under every potential entry word with a reference to the headword under which the idiom appears. *NTC's American English Learner's Dictionary* (*NTC*) includes such a list after its A-Z section, as does *CDAE*. In electronic dictionaries, finding idioms is no problem, since typing in the idiom in the prescribed way, usually within quotes, will generally take the user directly to it. In most print dictionaries, verbal idioms such as *land on your feet* are “run in” at the end of the entry for one of the key words of the phrase, in this instance *land* or *foot*. In dictionaries that give headword status to different senses, such as *CDAE*, the idiom falls within the entry giving the sense of the word under which the idiom appears, so *CDAE*'s idioms index shows this idiom appearing at *land* in the sense of “arrive.” This arrangement (or microstructure, to use the technical term) calls for a finer sense distinction of the components of the idiom than is necessary in other dictionaries, which can lump together idioms without regard to the meanings of their key words. For print dictionaries, the question of which word in the idiom is most likely to be sought by the user is still one that has to be addressed and that is sometimes impossible to answer. Should the idiom be placed under the first word, or the most important word? Sometimes the first word is variable, as in *shed* or *throw light on*. Sometimes it is not easy to say which word is more important, as in *hang fire*. Most dictionaries prefer to list idioms under the first word, though exceptions are common. Absolute consistency is purchased at the price of the reader's confusion and frustration. Idioms indexes can help a great deal if readers use them, but so far few dictionaries include them.

The problems of alphabetization in general monolingual dictionaries are minor compared to those in scientific and technical works, which usually have a much higher percentage of phrasal entries, often 50 percent or more of the total. Rather than try to alphabetize *nuclei nervi vestibularis* letter by letter, for example, medical dictionaries generally employ the main entry/subentry system (sometimes called *nesting*), alphabetizing under the canonical form of the governing noun and then alphabetizing (usually letter by letter) within the subentry field. The term quoted above is alphabetized under the entry *nucleus*. The fact that its form of *nucleus* is plural has no bearing on its placement amidst other subentries that begin with the singular form. Within the field of terms beginning with a form of *nucleus*, its alphabetic place is determined by the letters NERVIVESTIBULARIS. *Stedman's Medical Dictionary* lists the following words in this order:

- Monakow's nucleus
- motor nuclei
- motor nucleus of facial nerve
- nucleus motorius nervi trigemini
- motor nucleus of trigeminal nerve
- motor nucleus of trigeminus
- nucleus nervi abducentis

The letter-by-letter alphabetization of subentries is modified by ignoring some prepositions (such as *of* in the examples above), connectives, and articles.

The treatment of chemical terms is particularly troublesome, because systematic chemical names (like *methylcellulose*) are formed simply by sticking together the parts needed to describe the terms chemically, and it is a matter of indifference whether they are written as one word or two. Should *methyl cellulose* be alphabetized under *methyl* or *cellulose*? *Stedman's Medical Dictionary* partially resolves this problem by making an exception to its usual entry/subentry method of alphabetization. It lists most *methyl-* terms, for instance, as headwords, whether one word such as *methylprednisolone* or two words such as *methyl red*. But it cannot list all *methyl-* terms as main entries. *Active methyl*, *methyl alcohol*, and a number of other terms still appear as subentries under *methyl*. In dictionaries, one soon finds, if there is a possibility that certain forms may occur that do not fit into any category, one can be sure that they will occur. The best advice I can give is to set policy only after having had some experience with the kinds of entries to be encountered. Having set it, make minor adjustments or major exceptions when necessary, but once the project is well under

way, don't make any major systematic changes. The programming that would have to be done at that point, given the complexity of the rules of alphabetization in a work of this sort, would probably require so many adjustments before all the bugs were out that it would not be worth doing, and might well introduce problems where there were none formerly. One is better off sticking with a less-than-perfect system than risking disaster.

Another major problem of alphabetization in technical dictionaries is the ordering of non-English alphabetic characters; of different typefaces of the same letter; of the same letter with superior or inferior figures, letters, or symbols; of entries consisting wholly or partly of numbers; and of non-alphabetic symbols. Practice in these cases varies from one dictionary to the next and, all too often, within the same dictionary. The issue must be addressed in sufficient detail in the dictionary style manual, to be discussed in Chapter 7. (See p. 363.) The particular order in which the entries are listed is of little importance so long as the problem of determining an order of entries is appreciated early in the project and a start is made at elaborating a fixed sequence. Almost surely, new terms will have to be added to the list in the course of compiling the dictionary; but once the order of entries is established, it should not be changed mid-course. For example, consider the possible ways in which the following entries could be sequenced:

Hg	hg ⁻¹
hG	hg _∞
HG	√hg
hG	HG
lhg	h _g
hgl	

This list is by no means far-fetched. In fact, it is simple compared to many technical glossaries.

ENTRY COUNTING

We must now approach the delicate subject of what is meant by saying that a dictionary has 50,000 or 100,000 or 180,000 entries. The current situation is clouded by the deliberately confusing nomenclature used by many dictionary publishers, eager to tout their books as bigger than their competitors'. They refer to "references" without specifying what they mean. In some cases, "references" does refer to entries, in other cases, clearly not. Sometimes claims are made about the number of definitions rather than entries – since there are many more definitions than entries

— in the hope that the user will not pay much attention to anything but the size of the number and will use this number to make comparison with another dictionary's entries. The tradition in British lexicography in the nineteenth century and for all but the last decade of the twentieth has been less competitive than in America, hence less dependent on publicizing invidious statistical comparisons, of which entry count is a favorite. During this period the entry count in British dictionaries was scarcely mentioned in publicity releases, and if it was, it usually referred unambiguously to the number of headwords. Unfortunately, British dictionaries have begun to play the numbers game of claiming more of *something* than anyone else. What is a dictionary entry?¹⁴

As developed in the competitive world of American dictionary publishing, entry counting is based on two presumptions:

1. Every word or phrase that is explicitly or implicitly defined, so long as it is clearly identifiable, is an entry. Typically such entries appear in boldface type.
2. The more entries one has, or can claim, the better.

The system is designed to maximize the number of entries one can claim, but it is neither illogical nor fraudulent so far as its rules are scrupulously followed. Unfortunately, they are not, and dictionaries have with disturbing frequency made claims that are hard to justify even by the rather liberal system I shall describe. I once had a critic take me to task for the small type of my dictionary and in the next breath praise the larger, more legible type of a competitor that claimed almost the same number of entries as my dictionary. Why, he wondered, could we not have produced a work of comparable legibility? The answer was that we had not lied about the number of entries, whereas our competitor claimed to have 30 percent more entries than his dictionary actually contained.

Here is a sample entry article from *Random House Webster's College Dictionary* (RHWCD):

par.a.chute (par'ə shōōt), *n.*, *v.*, **-chut.ed**, **-chut.ing**. —*n.* **1.** a folding, umbrellalike, fabric device with cords supporting a harness or straps for allowing a person, object, etc., to descend slowly from a height, esp. from an aircraft. —*v.t.* **2.** to drop or land (troops, supplies, etc.) by parachute. —*v.i.* **3.** to descend by parachute. [1775–85; < F, = *para-* PARA-² + *chute* fall; see CHUTE¹] **-par'a.chut'ist**, **par'a.chut'er**, *n.*

The following items are entries:

1. The main entry, or headword, i.e., *parachute* as a noun, the first part-of-speech specified.

2. Any other defined part-of-speech of the headword, i.e., *parachute* as a verb.
3. Inflected forms that are actually shown, such as *parachuted* and *parachuting*, even if represented in shortened form. Verbs having identical past tense and past participle forms in *-ed* and present participle forms in *-ing*, such as *rush*, usually do not show any inflected forms following the main entry. In that case, no other entries should be counted, even though the reader can surmise by their absence what the inflected forms are. Implication is not enough. A dictionary entry must actually appear in the article.
4. Run-on derivatives, such as *parachutist* and *parachuter*.

Thus, the article for *parachute* above includes six entries, not one. It is worth pausing to consider what does *not* count as an entry. Transitive and intransitive verb uses, such as defs. 2 and 3 of *parachute*, do not count as two entries. Individual definitions never count as entries. A word like *run* may take up a whole page and count for no more than two entries, one for each part-of-speech. Pronunciations and etymologies have no bearing on entry count. Lists of synonyms and synonym discussions do not count as entries, nor do usage notes.

The following items do count as entries, though they do not occur in the entry for *parachute*:

5. Idioms or other defined expressions “run in” an article, such as *at sixes and sevens* in the following entry for *six* from *Webster’s New World College Dictionary* (WNWCD):

six . . . **1** the cardinal number between five and seven; 6; VI . . . **—at sixes and sevens** [Informal] **1** in confusion or disorder **2** at odds; disagreeing

6. Run-in variant forms – sometimes called *hidden boldface entries* – that are an alternative or longer form of the entry word, as in the following example from WNWCD:

creeper . . . **1** a person, animal, or thing that creeps . . . **8** the lowest gear in a truck, as for use on steep grades: in full, **creeper gear**

7. Other variants, whether alternative spellings or lexical variants, as in the following examples from RHWCD:

oys’ter.catch’er or **oys’ter catch’er** *n.* any of several heavy-billed shorebirds . . .

NFD. or **Nfd.** or **Nfld.**, Newfoundland.

tea’ serv’ice, *n.* a set of chinaware, silver, etc., for preparing and serving hot beverages, esp. tea. Also called **tea’ set’**.

Abbreviations provide a plethora of entries in very little space; thus *RHWCD* earns three entries for the abbreviations for Newfoundland. Geographical entries, which abound in most dictionaries, have many variants; biographical entries are usually short and are also a fertile field for adding to one's entry count. Even when such lists appear in a separate section following the A-Z section, they count as entries.

8. Words in lists following an entry for a common prefix, such as *un-*.

Each of these words is a valid entry only if two conditions are met: first, that the prefix is defined in the sense in which it is used in the word; second, that the word sans prefix (in this case, the affirmative sense without *un-*) is defined in the dictionary. If these two conditions are met, the sense of the word can be surmised. Good dictionary practice also requires that if a word has any additional meanings not covered by the basic form that is defined, it should not be entered in a list, although this is not a condition for entry status. For example, the usual meaning of *unexceptional* is "ordinary," although it can also mean "not exceptional." Therefore, it would be misleading to enter *unexceptional* in a list of *un-* words even if *exceptional* were entered and defined.

Lists of this sort may be found in many dictionaries under *non-* and *re-* as well as *un-*. Although sometimes ridiculously inflated to include terms that seldom if ever occur, in principle such lists are legitimate. Because such prefixes can be applied to so many words and because the meanings are clear, including entries for all of them would be a waste of space. Many people believe wrongly and naively that if a word does not appear in a dictionary, it does not exist and may not be used. Lists may serve the purpose of comforting such people by including many forms that the compilers could not otherwise include. The same arguments, by the way, apply to the inclusion of run-on derivatives. In run-on derivatives the stress pattern, which is shown, often differs from that of the headword. For these reasons, both run-on derivatives and lists are useful space-saving devices if used with restraint, and it is perfectly proper to count them as entries.

The main entry form in a dictionary serves a number of different purposes:

- It indicates the preferred spelling.
- It indicates the usual printed form of the lexical unit, i.e., whether capitalized or not; whether considered foreign (and italicized) or naturalized.

- In the past, syllabication was routinely indicated with a centered bold-face period, a light vertical rule, or by some other symbol. Some dictionaries still provide this information, but more and more are abandoning it. In truth, syllabication is much less useful now than in former years. Most texts are composed on a computer using wrap-around technology that avoids the need for a carriage return, as in the old-fashioned typewriter. One therefore rarely has to hyphenate text, except in uncommonly long words where the division is obvious. Moreover, when syllabication is given in dictionaries, it is often misleading. The purpose of syllabication is to tell the reader where a word can be divided. Why, then, do some dictionaries – even some ESL dictionaries addressed to non-native speakers of English – persist in syllabifying **an.y** or **man.y**, which should never be divided? Showing syllabication is not necessary for pronunciation, which is separately indicated. Inertia and inattention are the only answers. Nobody has paid the matter any notice in years. In my view, syllabication is no longer necessary and takes up space that can be better used in other ways. For foreign learners, especially, it can be confusing, because a word division within a word can be mistaken for a word space, and the reader sees the entry as two words when it is in fact one. Multiword lexical units, where there are actual word spaces, are particularly likely to be misunderstood. In ESL dictionaries, the one exception where syllabication might still be useful is in closed compounds such as **east.bound** or **school.mate**, because division here reveals the composition of the word, which might not otherwise be obvious to foreign learners.

Almost every criticism made of dictionaries comes down at bottom to the lexicographer's need to save space. The elements of style that so baffle and infuriate some readers are not maintained for playful or malicious reasons or from the factotum's unthinking observance of traditional practice. They save space. Every decision lexicographers make affects the proportion of space their dictionaries will allot to each component. It is perfectly fair for critics to question these decisions, but they must realize that the length of a dictionary is finite, and as large as it may appear to them, it is never large enough for lexicographers.

In most American college dictionaries, less than half of the total number of entries are headwords. A ratio of close to 50 percent is high, demonstrating that the dictionary does not have an excessive number of run-on derivatives, lists, or other entries with implied meanings. For many years, *WNW* maintained a policy of having no run-on derivatives,

but the smaller entry count this policy assured eventually compelled the editors to abandon it for competitive reasons.

It cannot be assumed that smaller dictionaries necessarily have lower ratios, since they also have fewer total entries, and there are limits to how many words can be run on as derivatives or entered in word lists. Their ratios are apt to be similar to those of college dictionaries and in some instances higher, depending on the policy of including derivatives. School dictionaries have far fewer derivatives, hence higher ratios, than adult dictionaries. ESL dictionaries, which are less concerned with total number of entries than native-speaker dictionaries, include no run-on derivatives or very few. Among ESL dictionaries, coverage of idioms, phrasal verbs, and collocations is more important than beating the competition in number of entries. This is one reason why even large ESL dictionaries contain on the whole fewer entries than native-speaker dictionaries of comparable size. What they include, they generally define.

GRAMMATICAL INFORMATION

Grammatical information is more essential for the person who is trying to speak or understand a foreign language than for the native speaker. It is not surprising, then, that ESL dictionaries should provide considerably more grammatical help than dictionaries for other audiences. In the 1960s and '70s, ESL dictionaries – then all British – gave very extensive and detailed grammatical information, particularly about verb patterns. The complexity of some of these systems baffled many readers, and the 1980s saw a trend towards simplification of grammatical coding, with the notable exception of the *Collins Cobuild English Dictionary*, which includes a wide outer margin on its pages for detailed grammatical coding and other information. Still, all the ESL dictionaries include a good deal of grammatical information. All distinguish between count nouns and mass nouns, for example, and give information about common verb patterns, the gradability of adjectives and adverbs, and the kinds of complements (such as *that* clauses) typically following certain verbs.

There is good reason to include such information in ESL dictionaries and in bilingual dictionaries, especially when the translation in the target language is a count noun whereas the source-language term is a mass noun, or vice versa. Some scholars have argued that it would be useful to include similar information in native-speaker dictionaries. New

grammars of the English language are being prepared with the help of language corpora that will change the way grammatical information is treated in dictionaries, beginning with ESL dictionaries but eventually affecting every type of dictionary.¹⁵ However, the needs and purposes of ESL dictionaries, which are essentially pedagogical, will always demand a more specialized treatment of grammar than in other dictionaries. Even the basic distinction between count and mass nouns is given to many exceptions among native speakers, who can and do pluralize nouns that ESL dictionaries call "mass." I have in my files "count" examples for the following mass nouns: "a tiresome *arrogance*," *Englishes* (now fairly common), *evidences*, *frightfulnesses*, *harms*, *impacts*, *knowledges*, *musics*, *nonsenses*, *poetries*, *politenesses*, *publics*, *resistances* (in nontechnical usage), "There is not one but many *silences*," *softnesses*, *surgeries* (in American English use), *unknowabilities*, *violences*. It is unlikely that even very large corpora will contain evidence of examples for all of these words, yet they exist. We must therefore be wary of presuming that practical guidelines for the foreign-born student have any theoretical basis or practical utility for the native speaker, who should not be constrained against using novel forms because of their rarity.

Context and variety of usage determine whether a noun can be made plural. In poetry, anything can happen. In scientific contexts, some words, like *etiology*, are often pluralized, though in ordinary usage they are generally considered noncountable. My experience as a dictionary maker convinces me that it is rash to suppose that any grammatical form cannot occur. Not the least value of lexicography is that one learns to be humble about one's own knowledge of the language.

General dictionaries provide basic grammatical information. Adult dictionaries indicate the part-of-speech of each entry, and the senses of the verbs are usually distinguished as to whether they are transitive or intransitive. Illustrative phrases serve to show the usual collocations or contexts in which each sense is used, thus providing a variety of grammatical information, such as whether a word takes an indirect object or whether it is usually used in the passive voice. Often other grammatical information is explicitly given. For example, the reader is advised that words like *economics* that are plural in form are construed as singular and that some words like *barrack* are usually used in the plural. Indication of the form of a word, especially capitalization, is also a common feature of dictionaries.

I can hear the reader protesting at this moment, "Doesn't the same argument you made about count versus mass nouns apply equally to any

grammatical information given in a dictionary? If a dictionary says, 'usually used in the plural,' isn't it being equally factitious for the native speaker?"

My response is that the two cases are dissimilar. When a dictionary says, "usually used in the plural," it may be assumed that the only evidence that the word was ever used in the singular is a single fifty-year-old, smudged and partially illegible citation slip of questionable authenticity. Its statistical incidence in a corpus search approaches zero. When a dictionary says "usually," one may assume it to mean "almost invariably." The count/mass distinction is hardly so reliable for native speakers. Secondly, all nouns can be distinguished by countability, whereas other kinds of grammatical information given by dictionaries (except for part-of-speech) are highly selective. They are not given at all unless the evidence in their support is overwhelming. Foreign learners need guidance in the mass/count distinction and should not be concerned about rare exceptions, but native speakers do not need such guidance, and I am afraid that descriptions of prevailing usages, even if based on overwhelming statistical predominance, will be taken – perhaps even conveyed – as if exceptional uses are errors when in fact they are simply deviations from the prevailing norm. In every endeavor one finds people who have violated all the formal rules of preparation for success, yet defy all predictions – and confound their more traditional competitors – by excelling. Remote probability is an avenue of discovery, and though often fruitless, such statistically invisible efforts may add up in the aggregate, as particles of matter in the vastness of the universe are said collectively to have more mass than all the heavenly bodies combined, to something grander than well attested generalization.

Beginning children's dictionaries often do not include part-of-speech indications, because such labels would serve only to confuse the child who had not learned the rudiments of grammar. Even intermediate children's dictionaries deemphasize part-of-speech labels. Scientific and technical dictionaries provide little grammatical information. Most do not indicate part-of-speech, although irregular plural forms, as for Latin terms, may be included. Since the study of grammatical usage of technical terms is spare or nonexistent, there is little basis on which to provide information or guidance. Given the encyclopedic nature of such works, grammatical data are of marginal relevance in any case.

In bilingual dictionaries, the grammatical categories of the source-language vocabulary and its corresponding translations should be consistent. "The reader has the right to expect," one linguist observes, "...

that if Japanese verbs are usually translated as English verbs, an adjective will not turn up without some compelling reason." If a Japanese adjective is translated as "is dark," for example, it would be misleading to translate a Japanese verb the same way; rather, it should be translated as "gets (becomes) dark, cloudy, etc."¹⁶

The use of the word *compelling* in the above quotation illustrates one of the most difficult grammatical problems for monolingual lexicographers, namely, how to decide when the present or past participle of a verb should be considered an adjective and entered as a canonical form in its own right. The case for *compelling* is compelling, but the status of many other words is less clear cut: *domed* (stadium), *marbled* (veneer), *tempered* (steel), *hardened* (criminal), *flourishing* (vegetation), *pleasing* (performance). There are thousands of such terms that seem to defy all attempts to place them in lexical categories by which to judge their fitness for entry status. Many, for example, do not have any meanings that are not already well covered by their underlying verbs. Because any verb participle can be used adjectivally in English and many nouns can be used attributively, the mere fact of occurrence with such grammatical functions obviously does not sustain the argument to include them as separate entries. But some participles are used adjectivally much more commonly than others, and by the use of grammatically-tagged corpora containing many millions of words, lexicographers can now compare the relative frequencies of participles used adjectivally. (See Ch. 6, p. 296, for more information on corpus use.) They can thus determine with a high degree of reliability which participles merit entry status as adjectives. They can also see whether some participles have slightly different shades of meaning when used adjectivally in some contexts than one could impute to them simply by knowing their verbal meanings. Judgment still plays a part, of course, in determining the status of participles, but corpus use has provided a more scientific basis for making the decision than was formerly available.

Dictionaries also provide grammatical information about *function words*, so called because they function primarily as grammatical devices, as for example to link two words or introduce a phrase, rather than as lexical units. *Of, for, at, and, but, and to* are examples of function words. In a study of the function word *of*, John Sinclair shows that with the use of a corpus, function words can be defined lexicogrammatically, that is, by showing their grammatical functions and lexical contexts in which they occur, as in *a piece of wood, a cup of tea*, etc. If one has enough examples of each type, one can categorize the types.¹⁷ Definition will be treated at length in the next chapter, but suffice it to say here that the

requirement for substitutability – that the definition for a word be substitutable for the word itself in most contexts – cannot be met with function words. Dictionaries have long realized that function words could not be defined strictly lexically, but have not quite known how to handle them. When *MW10* defines *of* as “– used as a function word to indicate a point of reckoning <north ~ the lake>,” it is employing a convenient style to get around the impossibility of stating what *of* means in “north of the lake” in a substitutable way. Other dictionaries employ different methods, but all end up trying to assign traditional definitions to them. Sinclair argues that the assignment of traditional meanings to function words is not helpful. Surely he is right, yet in practice it is not easy to describe meaning lexicogrammatically in a simple way, as we see in Sinclair’s own *Cobuild* dictionary, one of whose definitions of *of* is: “You use **of** after nouns referring to actions to specify the person or thing that is affected by the action or that performs the action.” To be sure, *Cobuild* then provides examples – “the kidnapping of the child,” “the arrival of the next train” – to show what it means, but so do more traditional dictionaries that give more traditional definitions. The fact is that with respect to function words, neither *Cobuild*’s lexicogrammatical definitions nor more traditional lexical definitions can be understood readily without examples. What *Cobuild* does, however, is to provide guidance to the user by showing the grammatical frames in which particular meanings occur.

PRONUNCIATION

The pronunciation system favored by native-speaker American dictionaries – unlike British dictionaries – is one in which the entry word is respelled in alphabetic characters with diacritical marks over certain vowels and with primary and secondary stress marks indicated. Called a *respelling system*, it represents the meaningful, distinctive sounds in a language but is not based on how the sounds are produced; that is, it is not phonetic. It is based on *phonemes*, the smallest units of sound that can be used to differentiate meaning. Phonemes are theoretical constructs, composites of similar but variously articulated speech sounds. Phonetic sounds, or *phones*, on the other hand, are actual speech sounds classified by the manner and place of articulation, that is, by the way in which air is forced through the mouth and shaped by tongue, teeth, palate, lips, and in some languages by the uvula. The *r* of *run* and of *fur* is articulated differently, thus phonetically different, but belongs to the same phoneme

in English and is represented by one symbol in most dictionaries. Likewise, the *p* of *pill* and *spill* and the *l* of *lean* and *feel* are not articulated identically, but are represented by the same phoneme. In some languages, the substitution of one of these sounds for the other in the same context would result in two different words; they are phonemically distinct in such languages, though not in English.¹⁸ In the speech of many American southerners, the vowel sound of *pen* and *pin* is in free variation; such speakers are unaware whether they use the high front vowel [ɪ] or the mid front vowel [ɛ] in either context, though in other contexts, such as those of *slept* and *slipped*, a distinction is observed. In contexts like *pen/pin* the difference is not phonemic to southerners, though it is to most other Americans, who interpret the two utterances as different words. If a dictionary were designed to represent only the American English of southerners, [ɪ] and [ɛ] in these contexts would be represented by the same symbol; but since American dictionaries seek to represent American English pronunciation in general use throughout the country, they use two different symbols, usually /i/ and /e/, to represent these sounds. Parallel situations exist in the dialects in many other regions. No phonemic pronunciation system can cope very well with this kind of dialectal variation.

The chief difficulty in representing pronunciation is that the letters of the alphabet often represent more than one sound, as *a* in *late*, *can*, *far*, and *care*, and sometimes two different letters represent the same sound, as *c* and *k* in *cool* and *kin*. Because the letter *a* is pronounced variously, different symbols must be used to represent its different sounds. That the distinction between the sound of *late* and the sound of *can* is phonemic can be demonstrated by comparing *can* and *cane*. A change in the vowel sound changes the meaning of the two words, which occur in the same environment. The vowel in *late* is usually represented in respelling systems by /ā/, that of *can* by /a/, that of *far* by /ä/, and that of *care* by /er/ or /â/. These sounds are then linked to "key words," familiar words in which these sounds can be distinguished. If readers see /ä/ in a pronunciation, they are supposed to look at the pronunciation key, usually printed on every two-page spread or in the front of the dictionary, and find /ä/ identified as the first vowel sound in *father*. The system works so long as everyone knows how to pronounce the key word and pronounces it the same way.

The most widely used alternative to a respelling system is that of the International Phonetic Alphabet (IPA), which is used in virtually all British monolingual dictionaries and in bilingual dictionaries. Because

IPA symbols are phonetically based, i.e., on the area and manner of articulation, they can be used either phonetically or broadly phonemically, that is, taking in a range of closely related articulations when these correspond to phonemes. The actual distinction between the IPA and respelling systems is that, though both may be used phonemically, the IPA is phonetically based, and respelling systems used in dictionaries are not. Since the sound of *late* is a diphthong – composed of two sounds – it is represented in the IPA by two symbols [eɪ], not one as in respelling systems. To avoid relying on the phonetic values of any one language, the IPA uses both alphabetic and nonalphabetic symbols, representing the sound of *can* by [æ], that of *far* by [ɑ], and that of *set* by [ɛ].

A system based on the manner of articulation can be used to produce the sounds of any language, even a language with which one is entirely unfamiliar. However, the notation to represent such sounds exactly must be much more elaborate than for a phonemic system. A purely phonetic system must indicate whether the tongue is high, mid, or low; front, mid, or back; whether a vowel sound is long or short, rounded, diphthongal, or retroflex (made with the tip of the tongue curled up toward the palate). In addition, the movement from one position to another, or glides, must be represented, stress pattern and pauses must be exactly indicated, and pitch or intonation may also be noted.

Because a purely phonetic description would be far too complicated for most dictionary users, all phonetically based systems used in general English dictionaries are broadly phonemic. For example, even though the *p* of *pill* and *spill* and the *l* of *lean* and *feel*, as I have noted above, are not phonetically identical, dictionaries whose systems are based on the IPA nevertheless use the same symbols – [p] and [l] – in each pair of words because the differences are not phonemically significant and distinguishing between them phonetically would serve no useful purpose. In technical language, they are *allophones* of the same phoneme. Foreign speakers will have to learn the slight differences in articulation through practice and by listening to native speakers, not through reading dictionary transcriptions.

Although a respelling system can work rather well among native speakers who are all familiar with the basic quality of English sounds, however they may differ in a few particularities, it does not work at all among foreign learners of English, who may not know how to pronounce *late* or *can* and who may bring into play their own foreign-language equivalents of these vowel sounds rather than use the English sounds. The phonemes in their own languages are not likely to be those

found in English. For ESL and bilingual dictionaries, then, it is obvious that a phonetically based system is necessary. In recent years, the IPA has been adopted by some American ESL dictionaries, notably *LDAE* and *CDAE*. But the American college dictionaries have so far resisted using the IPA (except for the introduction of the schwa [ə]), much to the consternation of linguistic scholars, and even a dictionary that boasts of its representation of "global English," the *Encarta World English Dictionary*, opts for a respelling system rather than the IPA for its pronunciations.

A panel of American lexicographers convened by the American Dialect Society to discuss this very issue offered a number of reasons for the reluctance of American dictionaries to use the IPA.¹⁹ Victoria Neufeldt, then of *Webster's New World Dictionary*, said that "the audience for a general dictionary [has] . . . an internalized, generally unconscious understanding of the relationship between English pronunciation patterns and the standard orthography," hence a respelling system is better. She also pointed out that the values of IPA symbols, which were based on the sound systems of the languages of Europe, would result in transcriptions like [fit] for *feet*, which most native speakers would find confusing since the transcriptions resemble other English words. Enid Pearsons of Random House agreed. Both maintained that pronunciations in general dictionaries should be kept simple and should limit the number of variants shown; an IPA system would be more complicated than necessary. Frederick Mish was not "optimistic about the chance of expanded use of the IPA in . . . [Merriam-Webster's] general dictionaries," as distinguished from learners' dictionaries. As a member of the panel, I offered the opinion that the expanding market for foreign learners' dictionaries would be the vehicle for growth by which the IPA might make inroads in the native-speaker dictionary market. ESL students, who are highly motivated to learn English pronunciations, are generally familiar with the IPA through their language courses. But in native-speaker lexicography there is a populist counter trend to make respelling systems even simpler, because studies have shown that the respelling systems currently in use are widely misunderstood. The simpler systems, which may be called "newspaper respellings," because they are commonly used to represent pronunciations in the popular press, would even abjure use of the schwa, substituting [uh] instead, and use [ee] for IPA [i] and [ah] for IPA [a].

The real question at issue is, "What's a dictionary for?" Does a dictionary for native speakers of the college size or smaller have a pedagogical or heuristic function as well as an information-giving function?

Should we tolerate some probability that a dictionary user will not understand everything perfectly in order to give reasonably accurate and exact information, or should we sacrifice that for a low level of basic comprehensibility for everyone? Is a dictionary essentially like a telephone directory, or is it more like a textbook? In a textbook, the reader often is challenged to take in more than can be easily assimilated. Learning does not come without a price. A directory poses no challenges but gives the merest amount of information. What is wrong with newspaper respellings? They rely on script – letters – rather than sound to convey sound. Respelling representation uses familiar letters but at least signifies through its use of diacritics over many that these symbols are not really letters. They are symbols for sounds. But when one renders *attention* as [uhtenshun] or *extravagant* as [ekstrahvuhgunt], one is using many different combinations of letters to represent the same sounds: sometimes *uh*, sometimes *u*, sometimes *ah*, sometimes *a*, etc. In rendering pronunciation, ideally one wants to represent each sound with one and only one symbol. Newspaper respellings are back-of-the-envelope and are not consistent; the same respelling can refer to two or more different pronunciations. They are also often inaccurate because the spellings often unduly influence the rendering of the pronunciation. Sympathetic as I am with readers who have difficulty with dictionary respelling systems, I am not in favor of simplifying pronunciations any further. In fact, I would like to see the IPA, which as we have seen can be used phonemically, gradually replace the respelling systems.

The arguments made by the lexicographers quoted above in defense of their dictionaries' use of respelling systems have some merit, but one must bear in mind that what was not stated was that marketing considerations play a very large role in making dictionaries reluctant to make this sort of change. The American college dictionary market is highly competitive, and no publisher wants to introduce a feature that might make his dictionary harder to use. The lexicographers argued that even the current respelling systems are widely misunderstood. But surely one of the reasons pronunciation is so confusing is that there is no consistency from one dictionary to the next in how pronunciation is represented. Some symbols are more or less uniform, but others such as the vowel sound in *fear* or the final sound in *very* are not, and even minor differences such as the use or absence of the breve over a symbol (for example, *ă*) can be confusing. The IPA would provide that consistency, not only among English native-speaker dictionaries but in ESL and bilingual dictionaries and dictionaries of other languages. In the current

situation, one never knows whether the difference in rendering the pronunciation of the same word in two dictionaries using different respelling systems is an accidental artifact of the systems or whether it results from a different pronunciation. Apart from that, if people are already confused by the current systems, what have we to lose by introducing a new, better system? It isn't a case of "If it ain't broke, don't fix it." The system *is* broke. Although there are many IPA symbols, in practice, an adequate phonetically based system for a general English dictionary need not be much more complicated than a respelling system. The reluctance of American dictionary houses to use the IPA stems in part from the use of nonalphabetic characters, such as [ʌ] for the vowel sound in *hut* or [ɔ] for that of *law*, or the use of [ð], called the *eth*, and [θ], for the *th* sounds of *the* and *thread*, respectively. There really are not very many such characters; about a dozen IPA symbols would be needed for English dictionaries, and these, especially the consonantal symbols, could be introduced incrementally. A dictionary could proclaim the next year The Year of the Eth as part of a ten-year plan to convert to complete IPA representation. The IPA symbols for vowels would have to be introduced at the same time to avoid confusion with the existing alphabetic vowel symbols. Linked to a pronunciation key at the bottom of the dictionary page, the conversion process would not be an intolerable burden. There has been a great deal of publicity given to the "globalization" of English, but much of it is just talk. Here is something concrete that can be done to make the representation of English sounds understood throughout the world, and a clever publisher could make capital of this effort to promote his dictionary.

I would argue that as one learns to pronounce a foreign language little by little, through continuing use, increasing familiarity with the IPA will eventually lead to a recognition of a more exact phonetic value for each IPA symbol, even when it is used phonemically in a dictionary transcription. Although some IPA symbols may be redefined phonetically to embrace a degree of variation to avoid undue complexity, they will still have an essentially phonetic character and will be completely divorced from any alphabetic script – a fundamental distinction. Implicit in the arguments of the lexicographers in defense of respelling systems is the conviction that most users of their dictionaries are familiar only with English. This may once have been true, and may still be true to a great extent, but it is changing. It is bound to change. English will not always be the dominant language in the world, and in the age of the Internet and international business organizations, the knowledge of languages

other than English will be seen as an essential part of one's education, as it already is in most other countries of the world apart from the US. In addition, the dictionaries I have been styling "native-speaker" are not just for native-born speakers. Many immigrants who have no difficulty speaking or comprehending English use "native-speaker" dictionaries for the same reasons native speakers use them, and they would benefit from the use of the IPA. It is time that we stopped defining the audience of native-speaker dictionaries as monolingual people too dumb to understand the IPA.

It is by now obvious to the reader that I do not view dictionaries as comparable to telephone directories, and, while I would not call them textbooks either, I do feel that a dictionary has an obligation to enrich the understanding of the user when it is possible to do so without interfering with the basic purposes of the dictionary. Indeed, in other respects all the college dictionaries recognize this obligation, for example by providing etymological information, usage notes, and synonym discussions. These are not essential parts of a general dictionary's purpose, but they are useful and dictionaries quite rightly include them. They should give the same attention to accurate pronunciation.

In a sense the whole debate about the IPA and respelling systems has been rendered moot by advances in technology that have made possible the audible pronunciations of dictionary entries by using a CD or an Internet connection. Although the use of audible pronunciations is an entirely welcome and exciting development in lexicography, in general only one pronunciation for each lexical item is given, and it may not accurately represent the pronunciations of all speakers. For ESL dictionary users, who basically need one representative pronunciation, the absence of variants will be of little account, and audible pronunciations work well.

One of the major problems in pronunciation parallels that of establishing the written canonical form: what pronunciation will be regarded as preferred, if any? In Britain, where dialectal variations are great, one dialect is preferred over others. The Received Pronunciation (RP) of southern England, based on the careful speech of educated speakers, especially those who are products of the best public schools and of Oxford and Cambridge, is still the standard. No general dictionary could possibly do justice to all the varieties in Britain. In the United States, on the other hand, there are fewer regional differences, and they are relatively minor. The most widely used pronunciation – whether it is acknowledged to be a standard or not – is the measure of priority. Perhaps the major national difference is that of *r*-pronouncers versus

r-less speakers (in words like *sore* and *paper*). In American dictionaries, the form of the *r*-pronouncers is usually given priority, even though the front-matter guide may acknowledge that *r*-less speakers also exist in large numbers and are not speaking incorrectly. "I wonder," says James R. Hulbert, "how many users of our dictionaries realize that when they see [and I would add, "when they hear"] *fārthar*, they are to understand that in the case of Virginians it means *fātha*?"²⁰

There is no question that widespread variations such as the alternative ways of saying *either* with initial /aɪ/ or /i/ must be represented. What to do about *pen* and *pin*, or *father* with or without the /r/, is another matter. As one pronunciation expert acknowledges, rules for converting one symbol to another based on instructions in the front-matter guide are difficult to follow: "Moreover, the dictionary user requires information to be explicit at the point of entry and rarely consults the good advice given in the introductory sections, which are usually read only by other lexicographers."²¹

The alternative is to show each variation in the pronunciation of each word, the practice generally followed by *NID3*. A number of prominent lexicographers have criticized *NID3*'s pronunciations as too elaborate and cumbersome.²² Some of the pronunciations are undeniably complex and difficult to follow, but it seems to me that the only synchronic, unabridged dictionary in English should record fully the variant pronunciations documented by regional linguistic atlases and by its own phonological records. Many of the criticisms that attended the original publication of *NID3* in 1961 showed that even professional writers had difficulty understanding many of the features of the dictionary apart from pronunciation, especially the presentation of usage information. But when one is trying to represent the language with a high degree of detail, some complexity is unavoidable. For the same reason, the *OED* is not the easiest dictionary in the world to use either.

Bilingual dictionaries generally show only one pronunciation. "[T]he foreign learner," one scholar comments, "would be at a loss when offered more than one variant."²³ Here as in ESL lexicography, audible pronunciations are generally to be preferred to written ones, but when transcriptions must be used, few variants should be included. Variant pronunciations are also kept to a minimum in school dictionaries. The student, whether native-speaking child or adult foreigner, needs direction. Both types of dictionaries are necessarily normative compared to general, adult, monolingual dictionaries.

Some scientific and technical dictionaries provide pronunciations and some do not. Because many technical terms are seldom pronounced outside of the lecture hall, it is often difficult to ascertain their pronunciations. Many pronunciations for scientific words are necessarily devised on the basis of analogy with other similar words whose pronunciation is known, since terms like *coccidioidomycosis*, for example, do not turn up frequently in conversation. The scientific vocabulary is preeminently written; numerous terms encountered frequently in print are rarely uttered. Thus, scientific and technical dictionaries necessarily adopt a normative approach to their pronunciations. Where a record of usage does exist, the pronunciations based on usage are given preference to those based on analogy, but in a large technical dictionary that situation is the exception rather than the rule.

James R. Hulbert remarks that dictionaries are less satisfactory in pronunciation than in spelling, meaning, or etymology.²⁴ The record of the spoken language is difficult to acquire, difficult to transcribe accurately and unambiguously, difficult to represent understandably in a dictionary transcription. Pronunciation remains a low priority in monolingual, commercial lexicography because lexicographers perceive that pronunciation is of marginal interest to their audience. Perhaps reflecting this state of affairs, one recent British dictionary, the *New Oxford Dictionary of English* (*NODE*), does not even give pronunciations for entry words it regards as too simple to warrant them, a policy that opens the door to the argument that simple words or definitions, too, might be omitted. But no one is ready to turn back the clock to re-embrace the "hard words" tradition of the eighteenth century.²⁵ The only comprehensive American pronouncing dictionary, John S. Kenyon and Thomas A. Knott's *A Pronouncing Dictionary of American English*, was published more than half a century ago (1944) and has never been revised. *NID3*'s full coverage of pronunciation may be interpreted as a gesture of atonement for its failure to revise Kenyon and Knott (published by G. & C. Merriam), a decision no doubt based on commercial considerations. For current pronunciations of American English, the reader has no choice but to consult general dictionaries. In ESL lexicography, the *English Pronouncing Dictionary* originally edited by Daniel Jones and revised by Peter Roach and James Hartman is the only general pronunciation dictionary in which both British and American English are given (almost) equal status, British English still being the default variety. The *Longman Pronunciation Dictionary* edited by John Wells covers mainly British

English, though some attention is given to American English. Both dictionaries use the IPA. These books, valuable as they are, are most at risk of being replaced by CDs in which pronunciations are given audibly.

ETYMOLOGY

Whenever the subject at a dinner or cocktail party among fresh acquaintances turns to what each of us does for a living, and I am introduced as a dictionary maker, inevitably the first response from my associates is that they find dictionaries fascinating, and in particular the derivation of words. From the number of cocktail party acquaintances who have in all seriousness, but with what sobriety I cannot say, affirmed that they loved to read – “yes, read” – dictionaries, one would suppose the nation to be studded with pockets of people neglecting their work and families to read dictionaries, especially etymologies. Somehow, I doubt it. I have never heard anyone say that he or she loved to read pronunciations. Although meaning is certainly most often sought by dictionary users, it does not have the cachet of etymology, which combines knowledge of other languages, especially the classical tongues, with arcane scholarship. It may come as a surprise to some readers, then, to learn that of all the elements included in modern dictionaries etymology has the least to do with the essential purpose of a synchronic dictionary. Etymology may be valuable in its own right, but it tells us little about current meaning and is in fact often misleading. Here is what a textbook on the development of the English language has to say on the subject:

The belief is widespread, even by some quite learned people, that the way to find out what a word means is to find out what it previously meant – or, preferably, if it were possible to do so, what it originally meant . . . [S]uch an appeal to etymology to determine present meaning is as unreliable as would be an appeal to spelling to determine modern pronunciation. Change of meaning – semantic change, as it is called – may, and frequently does, alter the so-called etymological sense, which may have become altogether obsolete . . . [¶] Certain popular writers, overeager to display their learning, have asserted that words are misused when they depart from their etymological meanings. Thus Ambrose Bierce . . . declared that *dilapidated*, because of its ultimate derivation from Latin *lapis* ‘stone,’ could appropriately be used only of a stone structure.²⁶

By the same token, Pyles and Algeo observe, we should have to confine the use of *eradicate* to roots, since it is derived from Latin *radix* “root”; *calculation*, from Latin *calx* “stone,” to counting pebbles; *sinister* to leftists; and *giddy* to those divinely inspired, since *gid* is derived from *god*.

The error of associating current meaning with past etymological meaning is commonplace also in the sciences, particularly medicine, where devotion to tradition and a belief in the sanctity of Latin combine to reinforce the fallacy. As the medical lexicographer A. Manuila points out,

Many words, whether technical or in ordinary use, have departed so widely from their original meaning that the latter can be of interest only to medical historians and linguists. Thus, etymologically, *thalassaemia* means "sea in the blood," *gonorrhoea* "a flow of semen," *artery* "an air vessel," *asphyxia* "a weak or suppressed pulse," *autopsy* "self-seeing," *embryo* "a young animal," *diploma* "anything folded twice," and *anatomy* "a cutting through."²⁷

Nathan Bailey's *An Universal Etymological Dictionary* of 1721 was the first English dictionary to treat etymology with consistent purpose and seriousness. Etymology was sometimes included in earlier dictionaries, but in a rudimentary way, often consisting simply of a language designation such as *L* for Latin. Bailey gave the Latin etymon and often more immediate sources, as in Italian or French. Though he too was inconsistent, Bailey is credited with having established etymology as "one of the requisites of any reputable dictionary."²⁸

How far we have come in our expectations of what general dictionaries should provide by way of etymology is illustrated by criticisms of recent dictionaries. Patrick Drysdale, a British lexicographer formerly active in Canada, lists the following etymological desiderata:

1. Source language or language family
2. First English form and/or immediate source
3. Date or period of entry into English
4. Changes in form and meaning in English
5. Intermediate stages
6. Ultimate known source
7. Semantic development
8. Ultimate underlying or hypothetical form, e.g., an Indo-European root
9. Cognates in related languages also derived from the underlying form
10. Other English words derived from the same base.²⁹

In the eighteenth century and the early part of the nineteenth, no one could have fulfilled Drysdale's standard for good etymological treatment, because the great advances in philology that were to transform it into a science had not occurred or were yet relatively unknown. What

are the principles that the famous philologists of the nineteenth century – Rasmus Rask, Franz Bopp, August Friedrich Pott, Jacob Grimm, Karl Verner, Ferdinand de Saussure, Karl Brugmann, and others – elaborated? Essentially, they discovered and proved that sound change was regular rather than hit or miss. Henry Cecil Wyld, in the introduction to his dictionary, *The Universal Dictionary of the English Language*, which gives exceptionally full attention to etymology, describes the discovery:

We soon learn to appreciate the truth of the great principle which guides all philological study, that a given sound is always changed, in one and the same dialect, in the same way, under the same conditions. Thus, for instance, in Greek initial *s* becomes the aspirate *h*, and between vowels is lost altogether; in Celtic initial *p* is lost; in Sanscrit, Persian, and the Baltic-Slavic families a sound which appears as *k* in Greek and Latin and as *h* in Germanic, develops into the sound of *sh* or *s*; Germanic has changed old *p*, *t*, *k* into *f*, *þ* ("thorn"), *h*, respectively, and so on. These changes take place everywhere, in all words containing these sounds, not in only isolated words here and there.³⁰

The importance of the last sentence above can scarcely be exaggerated, for up until this time etymologies had depended on chance similarities between words in different languages or on wildly speculative theories based on biblical stories to determine the derivation of words. As we have seen, Noah Webster devoted ten years of study to the world's languages with the aim of mastering the science of etymology, but since he was contemptuous of the principle of regularity of sound change and ignorant of the discoveries based upon it, his etymologies were worthless.

The modern etymologist must explain how every sound in the current canonical form of each word developed from older forms. If he cannot explain certain changes on the basis of systematic sound change according to established principles, he must have some other plausible explanation to account for them. Merely saying that the sounds of /d/ and /t/, for example, are similar will not do. Common sense, however, still plays an important role in good etymologizing. As the famous etymologist Walter Skeat declared, "observe chronology." "The word *surloin* or *sirloin*," he points out, "is often said to be derived from the fact that the *loin* was knighted as *Sir Loin* by Charles II, or (according to Richardson) by James I. Chronology makes short work of this statement; the word being in use long before James I was born." Borrowings are due to actual contact. "The history of a nation generally accounts for the constituent parts of its language. When an early English word is compared with Hebrew or Coptic, as used to be done in the *old* editions of Webster's dictionary, history is set at defiance . . ." ³¹

The importance of etymology for historical dictionaries is beyond dispute. The main purpose of a diachronic dictionary such as the *OED* is to trace the development of form and meaning over an extended period of time for each word in the language. In this sense every historical dictionary is a work of etymology, although it is not an etymological dictionary. An etymological dictionary shows a word's form and meaning when it first appeared in English, from what other languages the form may have derived, and often describes its intermediate and ultimate forms and meanings, but the meanings given are usually simple glosses used to distinguish the sense from past or future meanings. Its attention is directed to the evolution of the form, and in order to distinguish one form from another, meanings must sometimes be distinguished. The attention of a historical dictionary is directed to the evolution of meaning, and it must therefore show the various forms having the same meaning.

Synchronic dictionaries need not deal with etymology at all and some do not. Small paperback, general dictionaries usually omit it. Desk dictionaries include only very brief etymologies. American college dictionaries give more extensive etymologies. The British native-speaker dictionaries vary rather widely in their coverage, but the *New Oxford Dictionary of English (NODE)* stands out for giving more thorough etymologies than the others, and they are written in plain English without abbreviations. Many lexicographers have been skeptical about the public's appetite for dictionary etymologies. Clarence Barnhart states, with characteristic honesty, "Most dictionary users have little interest or use for etymologies."³² How does one square this with my acquaintance with untold numbers of people reading dictionaries and fascinated by the derivations of words? The people who think they love etymology regard etymology as stories – *interesting* stories – about word histories. Books recounting remarkable word histories – or histories that seem to be relevant to recent social or political events – have always been popular and successful with the public. In the nineteenth century, Dean Trench wrote a series of extremely successful and erudite works for a popular audience which, while not exclusively etymological, often included etymological information. In the early part of the twentieth century Ernest Weekley's books on etymology were very popular. H. L. Mencken's *The American Language* derived much of its popularity from its etymological information. Later in the century the works of William and Mary Morris and of Stuart Flexner drew wide and appreciative audiences, and more recently still William Safire's books have found a

wide audience. Such books naturally select just those words and expressions that will provide material for entertaining accounts and are often highly speculative. This is not meant in criticism – who would be churlish enough to disparage books that are both educational and fun to read? – but they are not comparable to the treatment of etymology in dictionaries, which must provide histories for many thousands of words, the great majority of which do not have charming anecdotes associated with their histories but which instead consist of an uninspiring list of foreign words and a handful of glosses. Nor can dictionaries indulge in clever speculation.

Children's dictionaries either omit etymologies entirely or include brief statements when the derivation is particularly interesting. For example, they may say that *iconoclast* comes from Greek words meaning "breaker of images." Subjecting children who are in the process of mastering their own written language to unfamiliar words in other languages does not make a great deal of sense. Etymologies are not included in ESL dictionaries or in bilingual dictionaries. To use space on this subject for a foreign learner, given the many other more profitable uses to which the space could be put, would be foolish, especially when dictionaries for native speakers cover the ground so much better than any ESL or "translation" dictionary could.

Etymologies are often included in scientific and technical dictionaries, especially those for the older sciences such as medicine, because they are mistakenly believed to be essential to an understanding of meaning. In fact, what is important in medicine is not a knowledge of Latin or Greek but of the modern meanings of combining forms derived from Latin and Greek. It certainly helps, if one has never encountered *nephroureterocystectomy* before, to be able to divide the word into its components – *nephro-*, *uretero-*, *cyst-*, and *-ectomy* – rather than perceiving it as an alphabetic blur. But to know that *nephro-* has to do with the kidney implies no knowledge of the Greek *nephros* underlying it, and a knowledge of the underlying Greek or Latin can be misleading, as we have already seen, by implying a relationship that no longer obtains. If one understands the limitations of etymology in conveying modern meaning, it is surely of value to know how the ancients perceived conditions such as *thalassemia* and *gonorrhea* and may even, in rare cases, throw light on some aspect of modern meaning. Etymology's chief value lay in its historic and linguistic interest, not in its relevance to the modern meanings of scientific terms. However, curiosity about etymology is widely held to be a sign of keen intelligence. It suggests interest in, if not familiarity with, study of

the classical languages, and Latin and Greek strike awe in the hearts of most people. Thus, by giving considerable attention to etymology, the compilers identify their dictionary with the prestige of the classical tradition and with a species of scholarship regarded reverentially by many potential customers.

Are etymologies useful to anyone? If they do not contribute to an understanding of current meaning, shouldn't they be omitted altogether from non-historical, general dictionaries? It is a question worth pondering. Patrick Drysdale considers the question and cites three reasons for including etymologies:

1. to provide raw material for the scholar and the student of the history of the English language;
2. to increase understanding of, and stimulate interest in, both language in general and one language in particular;
3. since a dictionary is a record of the culture of those who speak the language it describes, to provide clues to the history of the culture and its relationships to others.³³

Without etymologies, the entries in a dictionary appear to have been granted their present shapes by divine right, without relation to any other language and without relation to the past. To divorce a language from its past misses the opportunity to show language in its context, even though particular words may be very well defined without etymology. It is the difference between seeing a lion on the African plains and seeing it in a zoo. The zoo may afford a better look at the lion – but a much inferior view of lions. I come back to my view that a dictionary should be more than a telephone directory in which you search for a datum and find it. A dictionary should give one the opportunity to expand one's view of language by offering more than is necessary, even if by doing so it includes some material that some readers will have difficulty understanding. A book, incidentally, performs this function better than an electronic database because a book is easier to scroll when one doesn't know exactly what one is looking for, in the marvelously efficient process of data retrieval called turning the pages. And I do mean "exactly." Who hasn't been frustrated in a computer search by not realizing, after fifteen minutes of enduring brusque messages accusing one of complete and final failure, that the computer wanted one to type, for example, "N.Y." rather than "NY," or vice versa, without ever giving the slightest clue as to why one's original input had been rejected? I know it is easy for programmers to anticipate such variations, but the point is they don't, because programmers aren't editors, and people who edit books are.

Etymologies demonstrate that language changes in form and meaning. Even the brief etymologies in synchronic dictionaries remind people of this process and sometimes give glimpses of the way other cultures, or our own in times past, viewed particular words. The processes of semantic change, coinages of new words, and linguistic borrowing from other languages (*loan words*) have been thoroughly explored and discussed in fascinating detail in a number of studies.³⁴ Dictionaries have sought ways to exploit the public's broad interest in new words and meanings, but have so far failed to come up with a successful formula. Conventional dictionary treatment of etymologies does not usually shed much light on semantic development, for two reasons: first, because often there has been too little scholarly research and there is uncertainty or ignorance about the subject; and secondly, because dictionaries have better uses for the space and have usually compressed their etymologies as much as possible. In this respect *NODE* is exceptional in spending an unusual amount of space on etymologies, and perhaps its treatment will inspire other dictionaries to take a fresh look at presenting etymologies in a more readable way for the general user. However, etymological scholars seem anything but sanguine about the future of etymology as a discipline or as a feature in general dictionaries. Drysdale remarks, "In recent times, the emphasis on descriptive linguistics has resulted in a downgrading of the importance of the history of words, while the wider availability of etymological dictionaries has made a full statement of etymology in the general dictionary seem less important."³⁵ Yakov Malkiel, in a survey of the history of etymological research, laments, "With the rarest of exceptions, the best of our universities hesitate to offer at any level lecture courses or seminars on etymology," although they offer instruction in every other aspect of language study. Malkiel lists numerous reasons for this neglect and notes with evident regret the priority given dictionary-style compilations as opposed to scholarly monographs.³⁶ Etymologists engaged by synchronic dictionaries, and by most historical dictionaries for that matter, do very little original scholarly research but rely mainly on etymological dictionaries, which in turn rely on scholarly monographs and on other etymological dictionaries. Most dictionary etymologists, while often conversant with the classical tongues and modern languages, are not professionally equipped to do original scholarly research, and they wouldn't have the time to do it if they were. If the well of original research in etymology dries up, dictionaries will simply go on repeating past suppositions, right or wrong. Whereas we can expect the treatment of dictionary meaning to improve with corpus

use, and the representation of pronunciation to improve because of the recording and transcription of authentic speech, not to mention its audible presentation, the future of etymology is problematical.

In the eighteenth century and for most of the nineteenth, dictionaries of English placed the etymology near the beginning of the dictionary article, before the definition. This practice is still followed by the *OED* and other historical dictionaries, as well as by *NID3*, which presents its definitions in historical sequence. The Funk & Wagnalls unabridged *Standard Dictionary* of 1893 was the first dictionary to place the etymology after the definition, and since then the practice has been adopted by most other dictionaries. The reasoning was that since the definition is more often sought than etymology, it should appear earlier in the entry. The early placement of etymology within square brackets is seen as a kind of moat that the reader must leap across to reach his quest. I do not think it makes very much difference where the etymology is placed. If it is clearly marked as etymology, one loses no time sweeping ahead to its definition. Commercial dictionaries will, of course, go on declaring that they have shattered all precedent and opened communication to millions of frustrated dictionary users by their ingenious innovations, but often it's a matter of Tweedledum and Tweedledee.

SYNONYMS

One of Dean Trench's criticisms in his paper *On Some Deficiencies in Our English Dictionaries* (1857) was the scant attention given to "distinguishing of synonymous words." He calls for the discrimination of synonyms "likely to be confounded," such as *safe* and *secure*. The first dictionary to include a large number of such discussions was Joseph E. Worcester's great work of 1860, *A Dictionary of the English Language*, perhaps in heed of Trench's criticism. Worcester can be said to have established synonym discriminations as a standard feature of large dictionaries.

Dictionaries of synonyms

Dictionaries of synonym discriminations, variously known in the nineteenth century as *synonymicons* or *synonymies*, have a venerable history. The discussion of synonyms goes back to the ancients, and the earliest synonymies were of Greek and Latin during the Renaissance. French and German dictionaries of synonyms appeared in the eighteenth century. One of the earliest synonym dictionaries in English was John Trusler's

Difference between Words Esteemed Synonymous in the English Language (1766). Another very early synonym dictionary was that of Hester Lynch Piozzi, more widely known as Hester Lynch Thrale, the celebrated friend of Samuel Johnson. Her *British Synonymy: or an Attempt to regulate the Choice of Words in Familiar Conversation* appeared in 1794. William Taylor's small octavo *English Synonyms Discriminated* was published in 1813, and in 1816 George Crabb's *English Synonymes Explained*, the latter going through many editions. An edition of *Crabb's English Synonymes* was published in 1917, a century after the first edition.

The lack of any dictionary of synonyms in English before the end of the eighteenth century is not hard to explain. In the sixteenth and seventeenth centuries the extent of the entire English word stock was much smaller than it is now, and the problem then, as Taylor points out in the Preface of his book, was that the same word had to be used in a variety of contexts with different meanings. As the word stock increased, the lexicon absorbed words that were near synonyms of existing words, or words that evolved in meaning to become near synonyms. It was not until the word stock of English was sufficiently developed to embrace thousands of near synonyms that dictionaries of synonyms could be written with much effect.

As an example of the kind of discriminations given in such early books, here is part of Taylor's essay on *mirth* and *cheerfulness*, attributed to Addison: "Mirth is an effort, cheerfulness a habit of mind; mirth is transient, and cheerfulness permanent; mirth is like a flash of lightening, that glitters with momentary brilliance, cheerfulness is the day-light of the soul, which steeps it in perpetual serenity." Sometimes collocation was indicated, as in this essay on *brute* and *beast*: "Both these words are applied to animals, as distinguished from birds, insects, fishes, and man; but the term *brute* is confined to the untamed quadrupeds. [¶] We say, *beasts* of burden, never *brutes* of burden. The *beasts* of the field; the *brutes* of the forest."³⁷

At this point we must backtrack a bit and determine just what is meant by synonymy, in the sense of the state of being synonymous. Zgusta specifies three aspects of lexical meaning: the designatum, connotation, and range of application. *Designatum* refers to the essential properties of the thing or concept that define it; *connotation* refers to associated features; and *range of application* refers to the variety of contexts in which the word may be used.³⁸ Zgusta defines absolute synonymy as occurring when two terms correspond in all three aspects of designatum, connotation, and range. Absolute synonymy is rare among general terms but common

among technical ones, especially in medicine. For example, *Creutzfeldt-Jakob disease*, a progressive degenerative brain disease, has at least ten absolute synonyms (which I prefer to call variants), including *Jakob-Creutzfeldt disease*, *Jakob's disease*, *Creutzfeldt-Jakob syndrome*, *Jones-Nevin syndrome*, and *spongiform encephalopathy*.

If the correspondence is exact in one or two of the three aspects but not in all, the two words are near synonyms. *Beast* and *brute* have the same designatum – an animal other than man – but are dissimilar both in connotation and in range of application. It is precisely the object of synonym discussions to distinguish the ways in which two words differ in one or more aspect, usually that of connotation, and often of range as well. Such discriminations are based on individual point of view and depend for their effect largely on quality of expression rather than cogency of argument. At the end of his discussion of *beast* and *brute*, Taylor adds of another synonymist: "Dr. Trusler gives an opposite account of these words." Today both words are probably used metaphorically of people more often than of animals; these uses would be considered separate senses from those applied to animals but they would certainly affect the connotations associated even with the literal senses.

In the second half of the twentieth century, books of synonym discriminations were not nearly as popular as thesauruses (described below), and since comparisons of synonyms are much harder to prepare, there are fewer of them than of thesauruses. Richard Whateley's *Selection of English Synonyms* appeared in 1851 and Charles John Smith's *Synonyms Discriminated: A Dictionary of Synonymous Words in the English Language* twenty years later. The second edition of Smith's dictionary (1890) is an impressive work of 870 double-column pages, including etymologies, many quotations from well-known writers illustrating the words discussed, and an index. Smith frequently compares three or more words, and even as many as eight. James C. Fernald's *English Synonyms and Antonyms* was published in 1896 and revised in 1914. Fernald's discussions often involve a dozen words and consist of a sentence about each. He is one of the few synonymists to pay any attention to antonyms, though he merely lists them alphabetically at the end of each essay. The book includes a full index and a question-and-answer section for teachers. An interesting and useful feature is the inclusion of notes "on the correct use of prepositions" with the verbs discriminated. The *Funk & Wagnalls Modern Guide to Synonyms and Related Words*, edited by S. I. Hayakawa and the Funk & Wagnalls dictionary staff, was first published in 1968. Apart from Funk & Wagnalls, which has passed from the scene and is now, one might say,

defunked, the only dictionary house with a tradition of producing books of synonym discriminations is Merriam-Webster. Its latest collection is *Merriam-Webster's Dictionary of Synonyms* (1993). Collections of this sort are so difficult to prepare that no one wants to begin a new book from scratch; therefore, every existing collection is revised, abridged, supplemented, and used in every way possible to provide the bases of "new" books. Even the *Modern Guide to Synonyms*, for example, written in the mid-1960s, drew upon Fernald's synonym essays as a point of departure, although, since they were fifty years old, they provided little help in the main, and writing the book was an enormously time-consuming enterprise.³⁹

The much more modest success of synonym discriminations, as compared with thesauruses, is not surprising as soon as one realizes that the former are not really reference works. They may discriminate a few thousand words in all, but even were they to include twice or three times that many, they would hardly be comprehensive enough to assure users that they will find the word they seek, as a dictionary or thesaurus does. Their appeal is therefore much more limited, and for such books to find an audience they must either be written on a very simple level indeed or assume a degree of familiarity with word usage and meaning that not every reader has. For the reader to appreciate the finer distinctions among synonyms, the grosser distinctions have to have been assimilated and there has to be a thirst for greater knowledge of subtle discrimination. Interest in such books is not likely to be strong unless they are packaged in a way that represents them as having direct practical utility, which of course isn't true.

ROGET'S THESAURUS

In speaking of synonym dictionaries, I have been referring to works devoted to brief essays in which synonyms were discriminated. In 1852, Peter Mark Roget's *Thesaurus of English Words and Phrases* was published. Roget was an English physician who had already made a name for himself with his studies in physiology. The *Thesaurus* was a different sort of book, with words arranged, to use Roget's own description,

not in alphabetical order as they are in a Dictionary, but according to the *ideas* which they express. The purpose of an ordinary dictionary is simply to explain the meaning of words . . . The object aimed at in the present undertaking is exactly the converse of this: namely, – The idea being given, to find the word, or words, by which that idea may be most fitly and aptly expressed. For this

purpose, the words and phrases of the language are here classed, not according to their sound or their orthography, but strictly according to their *signification*.

Roget elaborated six classes of categories: abstract relations; space; material world; intellect; volition; sentient and moral powers. Within these broad classifications he placed numerous subdivisions, with subdivisions further divided. He recognized that many words would fall into more than one category and that the placement of a word in one or another category was a difficult decision. Even the first edition included an index to help readers find the signification they were looking for, but the presumption then was that the elaborate classification of thoughts would usually suffice without resort to the index: "By the aid of the table [of classification of subjects into divisions], the reader will, with a little practice, readily discover the place which the particular topic he is in search of occupies in the series . . . [I]f, during the search, any doubt or difficulty should occur, recourse may be had to the copious alphabetical Index of words at the end of the volume . . ." However ingenious the hierarchy of concepts, it was pure fantasy to suppose that any conceptual arrangement of the vocabulary of English was natural to most native speakers. It cannot be doubted that few readers could have found the meaning sought without recourse to the index, which has been expanded greatly through the work's many editions. The most recent edition, *Roget's International Thesaurus* (1992), edited by Robert L. Chapman, straightforwardly instructs the reader to begin with the index.⁴⁰

Conceptually vs. alphabetically arranged thesauruses

So closely identified is *Roget* with this kind of synonym work that the name, now in the common domain, has been adopted by many other synonym dictionaries, some using a conceptual arrangement and others organized alphabetically with a list of synonyms following every main entry. Both kinds of works are now called *thesauruses*.⁴¹ The claim made by proponents of the conceptual variety that theirs is the only true thesaurus is specious. "Thesaurus" was used long before Roget adopted it for his work in its Greek sense of a storehouse or treasure, with the specific meaning of an exhaustive survey of words, as in Ainsworth's *Thesaurus Linguae Latinae Compendiarius* (1736), and it has retained this sense in English. It was perfectly suitable for Roget to describe his own work by the word *thesaurus*, but there is no historical or logical basis for arguing that his use of the term is the only correct one and that *thesau-*

rus may not, with equal regard for its history, be used to describe other kinds of synonym dictionaries, and indeed other kinds of dictionaries.

Roget is used in the titles of dozens of different synonym books. The use or non-use of *thesaurus* and *Roget* (exactly like the use of *Webster* among dictionaries) is worth millions of dollars to competing publishers. The real conflict is between those who would like to reserve these words for their own exclusive profit and those who fear that their failure to use them would place them at an enormous marketing disadvantage. The unfortunate fact is that most buyers of reference books know very little about the merits of competing works but rely chiefly on a familiarity with their names. The argument that *thesaurus* has a true sense is merely a smokescreen to conceal the battle for commercial profit.

Both the alphabetical and the conceptual thesaurus rely on an alphabetical listing, but the conceptual arrangement leads the reader from an index to clusters of words centered upon a congeries of related meanings. Most alphabetical thesauruses do not have an index, the presumption being that any word the reader might want a synonym for will appear as a headword. (*The Oxford Thesaurus*, as we shall see, is an exception, as it does have an index.) Both systems involve a great deal of repetition of any given word. In theory the conceptual approach offers a greater likelihood that the reader will find the word being sought, because the words for similar concepts will be found in adjacent or nearby sections. The alphabetical arrangement, on the other hand, is easier to use, since the reader can often find synonyms by looking in one place rather than two or more.

Among the alphabetical thesauruses are *Merriam-Webster's Collegiate Thesaurus* (1988); and *The Oxford Thesaurus* (2nd edn., 1997) and *The Oxford Thesaurus, American Edition* (1992), ed. Laurence Urdang. To my mind these illustrate the weaknesses and strengths of an alphabetical arrangement. *Merriam-Webster's Collegiate Thesaurus* is evidently a revision of the *Webster's Collegiate Thesaurus* of 1976, though there is no indication of this on the copyright page. It, like the 1976 edition, is declared to be "a wholly new book" in exactly the same words. Both editions have entries for words that few would want a synonym for, such as *immaleable*, *immedicable*, *imingle*, *immix*, *plummetless*, *prodigalize*, *proemial*, *profusive*, *programma*, and many others of similar rarity. Surely anyone for whom these words sprang to mind would need no thesaurus. It seems apparent that the entry list and synonyms were created from a database and that the work was prepared with insufficient editorial intervention. However, this thesaurus does seek to make finer distinctions than other thesauruses by

having categories for “related words” as well as synonyms, and for “contrasted words” as well as antonyms. In both editions of *The Oxford Thesaurus*, the text consists of an alphabetically arranged thesaurus and a long index in which every synonym is listed with references to the headwords under which it appears, or to the headword entry under which it is listed, or both. The references specify under which numbered sense the synonym appears – a necessity, as there are many numbered senses in the alphabetical section. Among alphabetical thesauruses, it is unique. Within the alphabetical section, every synonym that is itself a headword is marked with a small circle, so the reader can turn to that entry for further synonyms. The thesaurus also includes an illustrative sentence for every sense to help distinguish meaning and show typical context. *The Oxford Thesaurus* is certainly the most sophisticated of the current lot of alphabetical thesauruses.

The conceptual scheme is associated with extreme inclusiveness. Rarely used words, non-English words, names, obsolete and unidiomatic expressions, phrases that are not lexical units but pseudodefinitions: all are thrown in together along with common words without any apparent principle of selection. For example, in the fifth edition of *Roget's International Thesaurus* (1992) – one of the best of the conceptually arranged works – we find included under the subheading *orator*: “Demosthenes, Cicero, Franklin D. Roosevelt, Winston Churchill, William Jennings Bryan, Martin Luther King.” Why not Pericles and Billy Graham? Why not Daniel Webster or John C. Calhoun? When one starts to include types of things, where does one stop? Another conceptual thesaurus, *Bartlett's Roget's Thesaurus* (1996), includes within its conceptual categories of synonyms many lists of types of things: styles of art and architecture, sporting activities, heraldic terms, and a lengthy list of phobias in which I was disappointed to find no phobia for dictionaries, only one for words (*logophobia*). It also includes a sprinkling of quotations. I just don't know why *giant sloth* should be in a thesaurus, but it appears in this one under the category **oldness** and the subdivision *prehistoric animal*, and appears alongside other synonyms like *woolly mammoth*, *pterodactyl*, and *trilobite*. At times these works seem to be a potpourri of everything the compiler can think of. The governing principle seems to be that the more there is, the more likely it is that something in the collection will prove to be useful.

The alphabetical arrangement is usually associated with more selectivity, although there are some exceptions. But, generally speaking, names, rarely used words, and types of things are excluded. Because of their arrangement, alphabetical thesauruses can follow the sense break-

down of a dictionary. *The Doubleday Roget's Thesaurus in Dictionary Form* (1977), for example, was modeled after the sense breakdown of *The Doubleday Dictionary* (1975). In a work based on selectivity, the compilers must have in mind a clear perception of their intended audience. Though the level of difficulty of the synonyms may vary somewhat, the necessity of excluding some words imposes a limited range upon most alphabetical works that is lacking in most conceptually arranged works. Alphabetical thesauruses tend to be more homogeneous than conceptually arranged works. The governing principle is that the compilers know best what the user is likely to need.

Dictionary treatment of synonymy

Because general dictionaries are truly reference works, the inclusion of synonym discriminations makes a great deal of sense. When they are not included, they are not missed; when they are included, they are a welcome superfluity. Some usage notes, such as that for *uninterested* and *disinterested*, may masquerade as synonym discriminations, and it is these the user is most likely to seek and find. The words included in true synonym discriminations are unpredictable. Dictionaries sometimes combine discriminations with brief lists of synonyms keyed to specific definitions. Both adult and children's dictionaries include synonymies, but the children's version is apt to be more didactic in keeping with its pedagogical purpose. Among general dictionaries, the discriminations of *The Century Dictionary* (1889–91) are outstanding for their quality and for the sensitive choice of quotations illustrating each word discriminated. They are neither wordy nor too slight but give fair attention to each word and show typical contexts in addition to the quotations.

ESL and bilingual dictionaries do not include synonym discriminations, since such rarefied distinctions are quite beyond the skills of their users and altogether irrelevant to their purpose. Their aim is not to provide exquisite sensitivity to another language but to provide the means for communicating competently with a native speaker of another language. That goal is tough enough without worrying about the fine distinctions of connotation. Synonym discriminations are strictly for the native speaker.

The larger question of synonymy and bilingual lexicography is a profound one beyond the scope of this book. For example, two words in English may be near synonyms, such as *brute* and *beast*, but their respective translations into another language may be completely unrelated.

Contrariwise, two words in English that are semantically unrelated may correspond in translation to two near synonyms in another language. This is an illustration of anisomorphism, discussed by Zgusta at length in his treatment of the bilingual dictionary.⁴²

I have mentioned technical dictionaries and said they differ from other kinds of dictionaries in having numerous absolute synonyms. As the reader may have surmised, I am not entirely easy with this statement. Can we really call *Jakob-Creutzfeldt disease* a synonym for *Creutzfeldt-Jakob disease*? It is a mere variant, having close kinship with spelling variants. Even in the case of less similar forms, such as the *Jones-Nevin syndrome* and *spongiform encephalopathy*, the absoluteness of synonymy with *Creutzfeldt-Jakob disease* is of a different order than that of ordinary language. *Jones-Nevin syndrome* and *spongiform encephalopathy* were names given to this disease by medical scientists who subsequently discovered that the diseases they had described were identical to one previously described by Creutzfeldt and Jakob. In medicine, this happens all the time. Recall the initial skepticism in the early 1980s that the virus concurrently identified in France and in the United States as the cause of what we now call AIDS was the same virus. It was first called by different names in the two countries. In America it was known as *human T cell leukemia/lymphoma virus-3*, abbreviated *HTLV-3*. Only later, after much discussion, disagreement, and confusion, was the name *human immunodeficiency virus (HIV)* settled upon. Indeed, since the syndrome it caused manifested itself by many different symptoms, there was considerable doubt that HIV could be the cause for AIDS in all of its varieties, and a small but obstinate minority of scientists is still not convinced. No examination of the usage of these terms in their immediate linguistic contexts could have established their identity, as would be the case with synonyms used in ordinary language. Scientific and technical terms are *said* to be the same by investigators on the basis of their examination of the detailed data describing the circumstances that these terms are supposed to signify. Subsequent research may show that the conditions described are not in fact the same, in spite of their having been used as absolute synonyms, in which case the terms now called synonyms will have to be reclassified. Unlike the words used in ordinary language, scientific nomenclature depends upon a process of naming and on the external, physical circumstances or properties of the things named. I believe that different names for the same entity should be called variants; to call them synonyms masks their fundamental difference from ordinary language, in which usage determines synonymy.

ILLUSTRATIONS

Nathan Bailey's dictionaries in the early eighteenth century included a number of woodcut illustrations, but it was not until the publication of John Ogilvie's *Imperial Dictionary* in 1850 that a dictionary gave prominence to pictorial illustration. *The Century Dictionary*, based on Charles Annandale's revision of the Ogilvie dictionary, is justly famous for the quality and number of its illustrations. The *Century* used some of the Ogilvie illustrations, but many others were drawn expressly for it or for the *Century Magazine*. The dictionary was published by the Century Company, whose *Century Magazine* was prized for the excellence of its illustrations. Since the manager of the company's art department, W. Lewis Fraser, was put in charge of assembling the *Dictionary's* illustrations, it is not surprising that their quality is so high. As Michael Hancher has observed, the *Century Magazine* broke new ground in the art of illustration both technically and artistically, and the publishers brought this expertise to bear in illustrating the *Dictionary*. Most of its illustrations are wood engravings, improved by a process involving the use of photography, but the *Century Dictionary* also includes a few halftones, a surprisingly early use of this form of illustration in a dictionary. Moreover, as Hancher shows, the choice of subjects and the manner of representation were unusually original for its time. For example, the *Century* illustrates a horse *bucking* (for the verb *buck*) with an illustration after a drawing by Frederic Remington, the popular Western artist and sculptor, whose art had already appeared in the *Century Magazine*.⁴³ Even today, lexicographers view the illustration of a verb as a bit daring. The illustrations of animals and plants are exceptionally fine and are often initialed by the artists, one of whom was Ernest Thompson Seton. Seton's illustrations of birds are particularly notable. No dictionary since – in a period of more than a hundred years – has come close to matching the elegance and sophistication of the *Century's* art.

Nowadays in the United States one takes for granted that every general dictionary will be illustrated, but it has long been debated whether the space used for illustrations contributes much to the basic purpose of a dictionary. To what extent do pictures help the reader understand meaning? In Britain, at least, the jury is still out. None of the leading dictionaries for native speakers published by Oxford, Collins, or Chambers is illustrated, although the *Encarta World English Dictionary* (*EWED*) (1999) devotes a considerable amount of space to illustrations. To some observers it is obvious that concrete objects such as forms of

architecture, animals, plants, and many other things marked by a specific shape, such as geometric figures, are more easily grasped by means of illustration than by verbal descriptions.⁴⁴ As a generalization, this is certainly true; but it is misleading and naive to suppose that an illustration makes the definition superfluous. A picture is at best a representative example of the type of thing defined, yet it does not encompass anything approaching the full range of possibilities defined by the term it is supposed to illustrate. It thus performs graphically what an illustrative phrase or sentence does verbally. To say that a picture obviates the need for definition makes the mistake of substituting an example representing a class of things – in Zgusta's terms the *denotatum* – for the qualities that define the object (the *designatum*). The idea that simple words like *dog* and *cat* need no definition will be discussed in Ch. 4; I find it primitive, ill conceived, and vacuous. I should say that simple words need no pictures; they certainly do need definitions.

Zgusta sees the primary purpose of illustrations as that of depicting unusual or unfamiliar things. I agree. No need to illustrate *dog* or *cat*, but *gnu* or *capybara* is another story. "The pictures should not be over-specific," he adds, "but only general lest the user accept a feature only accidental to the picture as criterial to the designatum."⁴⁵ In other words, if one's picture of a gnu has a peculiarly long neck, we may wrongly suppose that all gnus have necks that long. This is good advice that is not always easy to observe in practice. If one is depicting a rare animal or plant altogether foreign to one's own culture, one must necessarily base the illustration on a limited number of available photographs or illustrations. One collects what source material that is available. Ideally, an expert in zoology should review the drawings to see whether they are accurate; in practice, it is difficult to find experts in every category depicted, or to entrust them if they are found with the ultimate responsibility for warranting that the art is accurate. Some experts, alas, are not conscientious; others, from a misplaced desire to be obliging, are not critical enough to be a reliable help. Frequently the editor must rely on his own judgment based on the best information available. Nonetheless, Zgusta's advice is important to keep in mind. Although there may be some comparatively short-necked giraffes, we ought not to use them as our model for *giraffe*.

This raises the question of whether drawings or photographs are more useful for dictionary illustration. Photographs are necessarily of unidealized individual things, whether zebras, geese, or medieval churches. Drawings may combine features of many individuals and thus

represent a composite distillation of elements regarded as typical. If a drawing is done well, it is usually more informative, with its details more readily apparent, than any photograph (or halftone, as it is called when processed for printing). On the other hand, photographs have – or had, before the age of digitized images – an undeniable attachment to reality. There was always, in viewing a photograph, a slight thrill of appreciation that was missing in viewing a drawing; we knew that the thing photographed was really there and not designed to illustrate anything. It was authentic. In the age of computerized images, however, photographs can be easily altered. Photographic tricks once confined to Hollywood moviemakers and touch-up artists are now available to everyone. The cachet of authenticity of the photograph has been fatally compromised. On the other hand, this very flexibility in image-making allows dictionary art directors to modify photographs to make them more generic and remove unwanted detail. (For information on the practical aspects of dictionary illustration, see Ch. 7, p. 388.)

Even so, halftones are more difficult to use than line drawings and often less satisfactory. For one thing, a photograph must be much larger than a drawing to convey the same sense of distance and space rendered in a skillfully executed drawing. A photograph of a group of people, for example, must be severely reduced in size for dictionary illustration; one feels one is examining a postage stamp. Before digitization, everything in a photograph had to be uniformly reduced, the important foreground as well as the unimportant background; that is no longer the case, but there are limits to how much juggling one can do with differential sizing without making the photo look absurd. In an artist's drawing, proportions can be doctored more subtly to highlight whatever one desires.

Dictionary illustrations can be useful, but they are most often used as a marketing feature to promote one's own dictionary at the expense of the competition. Until the publication of the *American Heritage Dictionary (AHD)* in 1969, line drawings were *à la mode*. The *AHD*, with thousands more illustrations than any other dictionary of its time, included both line drawings and halftones. Although some of the halftones were of poor quality – always a problem in large printings with relatively inexpensive paper – the overall effect was successful from a marketing point of view. *AHD4* (2000) has introduced full-color illustrations. No doubt these add to the appeal of the dictionary for many users, but it is doubtful that color adds much to the explanatory power of the illustrations. Dictionaries that give expansive coverage to encyclopedic items are more likely to use halftones – for example, for people or places – than

those that do not. Thus *EWED*, with thousands of encyclopedic entries, includes many photographs. The *American Heritage College Dictionary*, Third Edition (*AHCD3*) (1997), the third-generation offspring of *AHD*, has many more halftones than line drawings, many of which to my mind are either not necessary to illustrate meaning or, if necessary, not as effective as drawings would be. They provide visual interest to the page; one's eye is drawn immediately to them. They are easier to obtain than line drawings and cheaper if one factors in the staff time required to plan and commission drawings. *Webster's New World College Dictionary* (*WNWCD*) (4th edn., 1999) has joined the trend towards using halftones as well as drawings.

School dictionaries include many more pictures in comparison to the number of entries than do adult dictionaries, and nowadays the illustrations in all the leading children's dictionaries are in color. Very often adults seek a definition merely to confirm what they think a word means; this is less often true of children, who may have no notion of the word's meaning. For example, an adult may know that a dormer is a kind of window in a house, though may be unsure just what *dormer* means. A definition that reads "a small, roofed structure extending out from a sloping roof and containing a window" may therefore evoke in his mind a picture of what he has previously seen but may evoke nothing at all in the mind of a child who has no preexisting visual idea of such a structure. But if a picture is included, the child may then recognize a dormer as something he has seen, or least be able to identify one in the future.

ESL dictionaries can also profit from pictorial illustrations but, unfortunately, pictures must compete for space with other more essential elements, such as grammatical and usage information and collocational aids. Because ESL dictionaries are often used in formal instruction, they must be of portable size. Therefore, although pictures are desirable and some are usually included, there is not much space for them. Also, because of the variety of cultures for which ESL dictionaries are intended, one has to be careful not to include anything in one's illustrations that would be offensive in cultures whose traditions are very different from those of the dominant cultures in America or Britain. A picture showing any casual contact between a man and a woman, for example, would be offensive in some Muslim countries, where such behavior is considered fitting only in the privacy of the home.

Pictorial dictionaries such as *Dudens* (see Ch. 1, p. 19) used for bilingual translation make excellent use of illustrations. For the most part, only line drawings are used. It is amazing how much information is

packed into each page of the Duden picture dictionaries through the grouping of drawings and the detailed labeling of the various parts in each illustration. With the exception of the *Collins Cobuild English Dictionary*, which has no illustrations, the larger ESL dictionaries – the *Longman Dictionary of Contemporary English* (3rd edn., 1995), *Cambridge International Dictionary of English* (1995), and the *Oxford Advanced Learner's Dictionary* (OALD6) (6th edn., 2000) – all make occasional use of composite, schematic drawings with many labels to illustrate complex objects. OALD6 also includes a number of color illustrations of types of things – such as foods, clothing, and animals – in addition to its black-and-white illustrations.

Technical and scientific dictionaries can and often do use pictorial illustrations, as well as charts and tables, to supplement their text. To what extent the pictures are useful is a matter of debate. The difficulty is that technical dictionaries are often intended to serve users on a variety of levels, from student to advanced research scientist. Whereas the student may find the illustrations helpful, the more advanced user may regard them as intrusive or oversimple and wish that the space had been used for additional entries. Generally speaking, dictionaries intended for a broad market, such as one-volume medical dictionaries, are helped by illustration. Charts and tables are obviously useful in such works, although as a rule they should not be used to omit entries but rather as a convenient reorganization of included terms. Chemical elements may be included in a table for purposes of comparison, but they should also be included as alphabetic entries. However, in some cases the table can substitute for certain terms. For example, a table of poisonous snakes may make unnecessary the inclusion of each poisonous species so long as the genera are entered with cross-references to the table.

Just how important are illustrations in a dictionary? The answer depends on the intended audience. The more elementary or pedagogical the work is, the more useful are illustrations. Scholarly or historical dictionaries have no need of illustrations. Bilingual dictionaries seldom include them, because pictorial illustrations of words in the source language are a monolingual function. Illustrations of some words in L2 (the language that is foreign to the user) – which is in effect what the Duden picture dictionaries do – would be helpful, though to my knowledge it is not done, undoubtedly because editors would rather use the space to add entries or expand translational equivalents.

Other graphic elements, such as type size and page layout – at least as important as illustration – will be discussed in Ch. 7.

FRONT AND BACK MATTER

Up till now we have been discussing only the A–Z section of dictionaries and have been ignoring the introductory material and appendixes, called in the book trade *front matter* and *back matter*, respectively. These sections vary greatly in importance depending on the nature of the dictionary. In adult native-speaker dictionaries, a guide to the use of the dictionary is now considered essential, but it is astonishing how little help was given the reader in early dictionaries. By the time of Johnson, however, it was traditional to include an essay on English grammar and a history of the English language in a dictionary's front matter. The trend in recent years has been to cut back on front matter because it is widely believed among lexicographers that no one reads it. Some dictionaries still have front-matter articles on the history of the language or of dictionaries, on current usage, or on the different varieties of English used throughout the world, but these, when they exist at all, are shorter than front-matter articles in the past. In the age of channel-surfing and Internet browsing, no one believes that users of general dictionaries will have the patience to read substantial articles in small type. Gone are articles about grammar, pronunciation, and, with rare exceptions, etymology.

Although front-matter articles are seldom read by dictionary users, they are often regarded as important by reviewers, who, faced with the daunting task of examining a 1,600-page book with two columns per page, seize on the most conspicuous elements to read. It takes less time and effort to comment on the jacket blurb and front matter, and perhaps to look up one or two favorite neologisms, than to make an informed estimate of the value of the dictionary as a whole. Many a dictionary that has been prepared by dozens of people and that consists of 1 to 3 million words written and revised repeatedly over a period of five to ten years has been judged on the merit of one or two short essays composed by consultants whose contributions to the dictionary did not extend beyond their own essays. One cannot fault critics for reviewing a dictionary's front matter; anything in the book is fair game. Yet compared to the overall effort of producing a college dictionary, front-matter articles are of negligible importance. The attention given by reviewers, however, makes them of considerable commercial importance. Thus front-matter articles are often written by prominent scholars or educators in an attempt to establish the authority of the work and lend it prestige. In some cases these scholars have been associated intimately with the work,

as, for example, Kemp Malone was with the *American College Dictionary*; but as a rule the association is more distant, sometimes little more than window dressing.

Ironically, the guide to the use of the dictionary, which is the only part of the front matter of demonstrable practical importance and the only one most readers ever use, is almost always ignored by reviewers. In order to detect omissions or ambiguities in the guide, the reviewers would have to have spent some time studying the dictionary's content and style, and this they are disinclined to do. Most guides describe every part of the dictionary article: entry word, syllabication (if given), pronunciation, inflected forms, various kinds of labels, cross-references, variants, etymologies, synonyms, and usage notes. The purpose of the guide is to describe as clearly as possible all the kinds of information included in the dictionary, show the reader how to interpret the data given (i.e., how to read the dictionary's style), and provide clues for locating as quickly as possible particular items of information. To put the matter simply, the guide answers the questions, "what's in it?," "what does it mean?," and "how do I find it?" Often parts of several pages from the A–Z section are reprinted to highlight particular features, such as illustrative sentences or usage labels, which are shown in another color or in a shaded area (called a tint block) or bracketed, and are linked to captions that identify the part and refer to sections of the guide where the item is discussed. This is an excellent use of graphics to provide the reader with a simple and clear index to the guide.

In general the American college dictionaries include more back-matter material than their British counterparts, with the exception of the *Concise Oxford*, which has forty pages of appendixes on a variety of scientific and linguistic subjects. It was formerly the practice to include most encyclopedic material – especially biographical entries, geographical entries, and abbreviations – in separate sections in the dictionary's back matter, but the current fashion is to integrate all such entries within the main A–Z section. Only *Merriam-Webster's Collegiate Dictionary* (MW10) (1998) continues to maintain separate sections for these entries in its back matter. Among favored back-matter sections in general dictionaries are guides to writing and punctuation and a miscellany of encyclopedic material, such as a list of the US presidents, population statistics, scientific and technical data (chemical elements, weights and measures, signs and symbols, etc.). The *AHCD3* differs markedly from the pattern by including only a section on Indo-European word roots by Calvert Watkins, thus restoring a popular feature of the first edition of

AHD that had been cut from the second edition. In former times a list of US and Canadian colleges and universities was certain to be included in all the college dictionaries, but now, remarkably, it appears in none. College students are only part of the market for these dictionaries, and although the word *college* or *collegiate* remains a strong selling point, publishers have apparently concluded that lists of colleges and universities can be dispensed with.

In school dictionaries and ESL dictionaries, front matter is much more important. In both cases, student guides or workbooks that include guides to the dictionary may be published separately as booklets. For the child, the guide to the use of the dictionary must not assume any prior familiarity with dictionary use. A children's dictionary guide is really not just to one particular dictionary but a guide to dictionary use. It must therefore be more detailed, yet written in much simpler language than that of an adult dictionary. The task is by no means easy and should be undertaken only by a highly skilled writer familiar both with dictionaries and with school curricula used in teaching dictionary skills.

A children's dictionary guide begins at the beginning, with the alphabet, and instructs the reader how to look up a word and how to use the guide words at the top of the pages. It then more or less covers the same ground that adult dictionaries do, but with numerous examples taken from the A–Z section and sometimes with questions or exercises for the reader. Teachers' manuals may be useful (as they are also in ESL dictionary instruction) in helping the teacher use the front-matter guide or separate workbook for classroom instruction. Whether separate or part of the dictionary's front matter, the guide for children is usually printed in large, attractive type, often in full color to highlight features and invite attentive reading. Information about the history of the language, etymology, or usage, if included at all, is generally incorporated in the guide, which may run to fifty pages or more.

For the ESL dictionary user, the front-matter guide or separately printed students' workbook and guide is basic to the understanding and use of the A–Z material. Although ESL dictionaries no longer incorporate long, detailed sections on the grammatical patterns of verbs, grammatical verb patterns are often taught in ESL/EFL programs that include instruction in using the dictionary. The back matter of ESL dictionaries contains various linguistic aids specifically for the foreign learner: lists of irregular verbs, sections on word formation, advice on punctuation. Sometimes standard encyclopedic material similar to that found in native-speaker dictionaries is also included. The *Cambridge*

International Dictionary of English (CIDE) uniquely includes an index of phrases for discovering where to find multiword lexical units. Some ESL dictionaries that have controlled defining vocabularies list the words used in such vocabularies in the back matter.

Technical and scientific dictionaries usually do not provide full guides to their use. Technical dictionaries that include subentries, for example, may give the reader some guidance on how entries and subentries are alphabetized, but often it is inadequate and does not account for many exceptions. Such dictionaries generally give scant attention to pronunciation even if it is routinely included within the body of the book. Transcriptions of pronunciation may be necessary for competitive reasons but seem to be considered of slight interest to users of technical books. The back matter of scientific and technical dictionaries is often devoted to tables of technical data, such as (in medicine) a list of normal ranges of different substances in laboratory tests.

Some dictionaries, often called encyclopedic dictionaries, go much further afield than the ones discussed so far. They may include a glossary of mythology, a manual of usage, a selection of quotations, a chronology of World War II battles, a section of maps in full color, or whatever strikes the publisher's fancy as a feature that will attract interest and invite purchase. The tradition of including encyclopedic material in dictionaries is as old as lexicography and was one of the practices deplored by Dean Trench in his celebrated paper of 1857. Although separate encyclopedic sections have nothing to do with the dictionary proper, there are no compelling logical reasons for condemning them. Their chief effect on the dictionary is to make it heavier to carry and more expensive to buy. Those who find their dictionaries heavy enough as they are will not rejoice to find biographies of every US president and vice-president contributing to their bursitis.

There is, to be sure, a touch of fraud in many encyclopedic dictionaries because they usually promise more than they deliver. By attempting to give the reader a bit of everything, they do nothing well, and they encourage the misconception that theirs is a better dictionary because it includes all sorts of aimless irrelevancies. But we must be mindful that we are talking about degrees of irrelevancy, since even the most respectable of commercial dictionaries boasts of certain encyclopedic features. *MW10* proudly displays an index of charts and tables and other marginalia, a device used a half century earlier by the *American College Dictionary*. A glance at the jacket copy of the leading general dictionaries shows how important these peripheral features are to their successful

promotion. *AHD4* boasts of having 4,000 photographs, drawings, and maps. *NODE* announces in yellow type that it has 12,000 encyclopedic entries. *EWED* simply says it has thousands of illustrations and thousands of geographical and biographical entries. To be fair, they also make claims about their coverage of the lexicon, but nonlexical features are an attractive feature for many buyers.

Yet there is a difference between these dictionaries and encyclopedic dictionaries, which are generally sold at substantially higher prices directly to the consumer through promotional mailings, TV commercials, or magazine ads rather than through bookstores. Encyclopedic dictionaries appeal to a different and on the whole less sophisticated market, and their promoters have to have material for their sales pitch for the large flyers that are sent out by the hundreds of thousands to potential buyers. These are no doubt being supplemented by Internet ads which may in the long run replace direct mailings. (However, judging by the contents of my mail, direct mail solicitations are anything but dead.) Whatever the medium, promoters want features to itemize, especially those that seem to distinguish their dictionary from all others ever made. No subtleties or qualifications of any kind are allowed to dilute the impact of blunt assertion. The idea that a statement in such an ad could be true or false would strike most mass-market promoters as the purest whimsy. They would, like Saint-Exupéry's businessman counting the stars, urge one to be serious.

Definition

What do we mean when we say we have defined something? The question, an ancient one, has been addressed by Plato and Aristotle as well as by modern philosophers, especially logicians and semanticists.¹ Although the distinction is not always made, and when made not always observed, logical definition is not the same as lexical definition. Logical definition – Richard Robinson calls it real definition, because it attempts to analyze things in the real world, as distinguished from words – has been the chief preoccupation of philosophers. When Socrates explores the meaning of virtue or truth, he is not seeking to define the words *virtue* or *truth* but the concepts that underlie them and the way people interpret these concepts. Philosophers have also concerned themselves with lexical (or nominal) definition, the definition of words, the subject of this chapter.

The traditional rules of lexical definition, based on Aristotle's analysis, demand that the word defined (called in Latin the *definiendum*) be identified by *genus* and *differentia*. That is, the word must first be defined according to the class of things to which it belongs, and then distinguished from all other things within that class. Thus *child* is a person (genus) who is young or whose relation to another person is that of a son or daughter (differentia).

Among other rules sometimes promulgated for definition are that a definition be equivalent to or capture the essence of the thing defined, that the definiendum not be included in any form among the words used to define it (called the *definiens*), and that the definition be positive rather than negative. Philosophers are not ignorant of the nature of linguistics or dictionary making, but it is remarkable how little attention they pay to the users of dictionaries. By contrast, lexicographers – all of them – pay a great deal of attention to the needs of their readers. For lexicography is a craft, a way of doing something useful. It is not a theoretical exercise to increase the sum of human knowledge but practical work to