For my parents

The Oxford Guide to Etymology

Philip Durkin

OXFORD UNIVERSITY PRESS
2

What is a word? Which words need etymologies?

2.1 What are words? 34
2.2 How new words arise 43
2.3 Lexicalization 49
2.4 Examples of lexicalization 51
2.5 Apparent reversals of the process 56
2.6 Cranberry morphs 56
2.7 Which words need etymologies? 58

In chapter 1 we encountered some of the main characteristics of etymology, its aims, and some important features of its methodology. We considered some examples of change in word meaning and change in word form, and began to look at some of the mechanisms by which both of these occur. We will return to these topics in more detail later. In this chapter and the next we will take a closer look at the main objects of study in etymological research, words. In etymological dictionaries a ‘word’ stands at the head of each dictionary entry, and the status and selection of these words can seem to be a given. However, the identification of words as coherent entities for study raises a number of quite complex questions. Additionally, selection of which words to concentrate on is a far from trivial matter.

2.1 What are words?

2.1.1 Problems of definition

So far in this book I have taken the term ‘word’ rather for granted, as being a self-evident one which any reader will readily understand. The concept is very familiar to a non-specialist, and the term forms part of general vocabulary and so does not have to be learnt by beginners in linguistics, unlike phoneme, morpheme, etc. In literate societies lay conceptions of word boundaries (i.e. where one word ends and another begins) are often very much bound up with literacy and the rules of various writing systems, but there is also at least some evidence that non-literate speakers of languages with no written form also have intuitions about word boundaries, as do children who have not yet learned to read in literate societies.1 However, it is also notoriously difficult to define a ‘word’ in a way which makes sense consistently at all levels of linguistic analysis. Specialists in morphology and also in phonology often grapple with this particular problem, and a full discussion would take up much more space than is available to us here. The discussion that follows will be brief, and will focus on those aspects which most affect etymological research.2

2.1.1.1 Spelling A non-specialist from most modern literate societies who is asked what a word is will probably say that the words in a sentence are the things written with a space on either side. This definition is unsatisfactory for linguists for various reasons. Firstly, not all languages have a written form, and even when they do they do not necessarily separate words. Certainly, the way that many languages are written tells us something about writers’ intuitions about what constitute words, but a definition on this basis runs the risk of circularity, and is also detached from any analysis of linguistic structure: by this criterion, words are the things that people write as separate words (i.e. with spaces between them) because they perceive them as separate words (whatever that may mean).

Written language also tends to be rather inconsistent in its treatment of certain kinds of units. Any survey even of published written English will show very considerable variation in whether some combinations of two nouns are written as a solid, or with a hyphen, or with a space between the two elements. Thus lunchbox can also appear as either lunch box or lunch-box, and even dictionaries do not agree on which to list as a preferred spelling. We would have to resort to some very odd reasoning to argue that lunchbox is one word but lunch box is two: both have the same meaning and behave the same way syntactically, as does lunch-box, and in the spoken language the pronunciation is the same for all three. This leads to the

1 See further Bauer (2003) 57, Sapir (1921) 34–5.
2 For detailed discussion of most of the points in this section see e.g. Bauer (2003), which I have largely followed here, or (with some slightly different perspectives) Adams (2001: 2–5), Booij (2007: 281–94), Plag (2003: 4–9).
fairly obvious conclusion that we are looking at three different spellings of precisely the same linguistic unit.

2.1.1.2 Meaning One useful and conventional way of thinking about words as linguistic units is that a word is a linguistic sign which has both form and meaning. (We will come to the very important concept of the arbitrariness of this linguistic sign in chapter 4.) Linguistic meaning is expressed by the combination of units in a sentence. This might seem to give us a shortcut to a definition of a word: words are minimal units of meaning in a sentence. However, a little reflection will present us with some major problems. It is not always possible to infer the established, conventional or institutionalized meaning of phrases from their constituent words: consider idioms like it's raining cats and dogs (and see further section 2.1.5 below). There is also ample evidence that people often analyse the morphological composition of unfamiliar complex words as and when they hear them in order to interpret their meaning, and that they do this as part of their general competence as speakers of a language. For instance, if someone knows the word vinaceous 'of the colour of red wine' they are unlikely to have any more difficulty in understanding the derivative formations vinaceousness or vinacously than the phrase very vinaceous, although they will probably never have encountered these particular derivative words before. (Both words are extremely rare, and even a Google search shows only a couple of examples of each.)

2.1.1.3 Phonological criteria Phonological criteria can provide very useful evidence about word boundaries. In some languages, probably including proto-Germanic at one point in its history, stress regularly falls at the beginning of a word. (In proto-Germanic more accurately on the first syllable of a lexical root, rather than on prefixes.) In some other languages, such as modern English, each word has a particular syllable on which the main stress will normally fall if that word is stressed in a sentence (e.g. 'kindness, in'epitude, incon'solable); but this is not true of all languages.

Some phonological processes apply only at particular positions in a word. In the history of German a sound change occurred by which obstruents were devoiced when they occurred word-finally, but not when they occurred medially or initially, giving rise to a situation in modern German where e.g. Rat 'counsel' and Rad 'wheel' are homophonous in the nominative singular (both /raːt/) but not in inflected case forms in which an inflectional ending

follows the obstruent (e.g. genitive singular Rates /raːts/ 'of counsel' beside Rades /raːdas/ 'of a wheel'). Some phonological processes, especially vowel harmony, typically operate across syllable boundaries within a word, but not across word boundaries. (See for example section 7.2.4 on i-mutation in the history of English.) However, other processes do apply across word boundaries, such as the assimilatory devoicing in English /haːtuː/ as a realization of have to. This is usually called external sandhi, following the terminology of the ancient Sanskrit grammarians.

2.1.1.4 Morphological criteria A commonly cited morphological criterion is that words are uninterpretable units, although there are exceptions, as for instance when expletives are inserted in the middle of a word in English, e.g. absobloominglylute.

2.1.2 Problems of analysis

In addition to there being no generally accepted and completely satisfactory definition of what constitutes a word, there is also considerable scholarly disagreement about whether some particular linguistic units should be regarded as words or as phrases, i.e. syntactic combinations of more than one word. In English it is notoriously difficult to define what constitutes a compound and what constitutes a phrase. To begin with an unproblematic example, it would normally be accepted that blackbird is a compound, and a black bird is a noun phrase. blackbird has reference to a particular variety of bird, and if someone calls a crow a blackbird they will be using the English language in an idiosyncratic way that is unlikely to be understood by anyone else. However, if someone refers to a crow as a black bird, then they will be making a simple factual statement, and in grammatical terms we will analyse their utterance as a noun phrase showing bird as a head modified by the adjective black. Conversely, female and younger male blackbirds are mostly brown. Even white blackbirds sometimes occur, and they are still blackbirds, albeit uncharacteristic ones, although they are not black birds. However, if we try to extrapolate from this unproblematic example precisely what it is that distinguishes a compound from a phrase, we start to encounter some real difficulties:

3 For discussion of this particular phenomenon from a number of different theoretical standpoints see Lass (1984).
• *blackbird* has a meaning not predictable from its component parts, whereas *black bird* refers very predictably to any bird which is black. But many phrases and idioms also have unpredictable meanings.

• *blackbird* is written without any spaces, *black bird* is written with a space. But compare again *lunchbox*, *lunch-box*, *lunch box*.

• In some languages an adjective will show agreement with a noun in a phrase but will show a bare stem form in a compound, giving a clear morphological criterion for telling phrases from adjective-noun compounds, but this is not the case in other languages such as modern English.

• *blackbird* shows stress on the first element, while *black bird* shows stress on *bird*, the head of the phrase. But consider *blackcurrant*, in American English typically *black currant*, but in British English typically *black currant* (except sometimes as the first element in a compound, when the stress may be shifted, e.g. *black currant bush*). Consider also idiosyncratic cases, such as street names ending in *street* (e.g. *Downing Street*, *Coronation Street*, *Ship Street*) as opposed to those ending in *road*, *lane*, *avenue*, etc. (e.g. *Station Road*, *Cemetery Road*, *Park Lane*, *Shaftesbury Avenue*).  

This last point in particular is the subject of much debate, but it is sufficient for our purposes to know that there is as yet no clear consensus. In the case of adjective-noun compounds, gradability of the adjective can be a safer test, at least if the adjective is gradable:

• We may talk about *a very black bird*, or indeed a very black *blackbird*, but not *a very blackbird*.

However, this criterion often conflicts with what we might predict from the position of the stress. *red admiral*, the name of a type of butterfly, has stress on the second element, suggesting phrasal status, but we cannot speak of *a very red admiral* or the *reddest admiral* (at least, not if we are speaking about the butterfly; either phrase would be perfectly plausible if referring to the left-wing politics or the flushed face of a naval officer).

---

4 For a useful discussion of these see Plag (2005).

5 For a recent summary see Bauer (2006a), and also Bauer (1998a); for a sample of rather different views see Booij (2007) or Giegerich (2004).

---

2.1.3 Why these are not major problems for etymology

I have introduced these issues largely to show that the use of ‘word’ and ‘compound’ is not always uncontroversial, and because it is important to realize that the simple statement ‘etymologists study the origins of words’ may not really be so simple as it at first sounds.

Whatever definition of the term ‘word’ we adopt, etymologists cannot avoid interesting themselves very closely in many units much larger than the word. Very many phrases have complex meanings and complex histories which require etymological explanation. Furthermore, many single words have their origin in what is sometimes termed the univerbation of what were originally phrasal units consisting of more than one word, e.g.:

• *upon* < *up* and *on*

• *goodbye* shows a contraction of *God be with you*, with remodelling of the first element after *good day*, *good night*, etc.

• the phrase *at one* > the adverb *atone*, on which the abstract noun *atonement* is formed, which in turn gives rise to the verb *to atone*.

In some other languages, such as French, lexicalized phrases frequently occur in meanings which are typically realized by compounds in English, for instance French *sac à main* beside English *handbag*. We can also examine the etymologies of units smaller than the word, for example derivational affixes such as *pre-, un-, -ness*, etc., and even morphological inflections, although these do raise some rather different issues, which we will explore in chapter 4.

Conversely, if we are studying a contemporary language, or even a past stage which has a large corpus of surviving evidence, then we cannot possibly pay attention to the etymology of every word ever uttered, or even every word ever recorded, in that language, and nor would we want to. As we will see in section 2.2.4, the lexicon of every language is constantly open to new words, formed according to the productive word-forming patterns of that language. Nearly all such new words are immediately transparent in meaning (when heard in the appropriate context) to other speakers of that language. Additionally, nearly all such words fail to enter more extensive usage, and remain ‘one-offs’ or nonce formations (although the same word may well be formed again, quite separately, by other speakers on other occasions).
2.1.4 Word forms and word meanings

If we return to the expression of meaning by words, we can observe that some words, like a or the, have grammatical content but no other meaning content. Other words, like haddock or ankle, have clear meaning content. Many words have multiple established meanings, i.e. they are polysemous, and we can only tell which meaning is intended from the context of a particular utterance. For example, we can speak of someone working in an office (a physical place) or holding an elected office (an abstract social role), or we can say that a container is full (there is no room left in it) or that the moon is full (none of the side turned towards the earth is in shade). In fact, meanings are often stretched or extended in particular contexts. It is only when particular new or extended meanings of words in particular contexts become institutionalized, i.e. used fairly frequently by different speakers of a language, and perhaps extended to other contexts, that they begin to be recorded in dictionaries. We will return to this point and its importance for etymological research in chapter 8.

Additionally, we need to distinguish between different homonyms, i.e. quite separate words which happen to be identical in form. For instance, distinct homonyms are shown by file ‘type of metal tool’ (of Germanic origin) and file ‘set of documents’ (a borrowing from French). In this instance the words are distinct from a synchronic point of view, since there is no semantic common ground between the meanings which they realize, and also from a diachronic point of view, since they have different histories. However, these two criteria do not always coincide, as we will explore in detail in section 3.3.

Meaning is also expressed by the inflections of a word, e.g. in the singular/plural distinction between giraffe/giraffes, board/boards, fish/fishes, man/men, etc. Technically, these inflected forms are distinct word forms, which belong to a single unit called a lexeme. In order to identify the lexeme to which the word forms giraffe and giraffes both correspond, we normally use what is called the citation form, i.e. the form that we can look up in a dictionary. So giraffe is the citation form of the lexeme which has the word forms giraffe and giraffes (also giraffe’s, giraffes’). Sometimes small capitals are used to identify lexemes, e.g. GIRAFFE, MAN. Note that in the case of man/men the morphological relationship is realized by variation in the stem vowel, rather than by an inflectional affix (see further section 4.4.1).

Sometimes we find the phenomenon known as suppletion, where word forms of different historical origins stand in the same sort of relationship, within a grammatical paradigm, as inflected forms like giraffe and giraffes do to one another. Thus, was and is are not inflected forms of be (they are of a quite different historical origin), but they stand in the same paradigmatic relationship to it as opened and opens do to open. Similarly, worse and worst stand in the same paradigmatic relationship to bad as poorer and poorest do to poor. We can say that be, was, and is (and also are) are word forms of the lexeme BE, and that worse and worst are word forms of the lexeme BAD (and also of the lexeme BADLY). Interestingly, in the case of worse and worst this pattern is relatively modern. Both forms go back to the Old English period (Old English wyrs and wyrst), and they have been the antonyms of beter and best (Old English betra and betst) throughout their history in English, but the adjective in the general sense ‘bad’ to which they correspond (again suppletively) as comparative and superlative in Old English is yfel (modern English evil). In early Middle English we find a new adjective ill in many of the same senses as evil, and worse and worst are also found as its comparative and superlative. Finally, bad becomes increasingly common in senses formerly expressed by evil and ill, and gradually worse and worst become established as its comparative and superlative forms. However, there is a long transitional period in which worse and worst are found in paradigmatic relationships with all of these three words, e.g. we find examples of from evil to worse, from ill to worse, and from bad to worse. Thus patterns of suppletion can vary over time, and can also vary in the usage of particular individuals or speech communities within a particular period.

Suppletion is quite different from the phenomenon where different variants realize the same grammatical form of a single lexical item. Modern standardized written languages do much to disguise this sort of variation, but consider the regional differences in pronunciation between for example /tɔθ/ tooth in the English West Midlands as against /tʌθ/ elsewhere, or the variation in the pronunciation of either as /aiθ/ or /aɪθ/ in the speech of different individuals in both Britain and the US. This is an issue that we will look at in much more detail in chapter 3.

In this book, I will normally use ‘word’ rather loosely in the sense ‘lexeme’, and I will refer to words by their citation forms. This is not normally a problem in etymological work, so long as we have a more sophisticated terminology available for instances where we need to tease the various
distinctions apart more carefully, and so long as we remain aware of the bundle of different forms and meanings which a single word may show.

2.1.5 Idioms

As we have noted, units larger than a single word also often have conventional or institutionalized meaning which is not predictable from their component parts. Idioms are by their nature constructions which are stored in one’s memory and form part of one’s competence in speaking a particular language, even if this only involves selection of the correct preposition or adverb in verbal constructions such as to sober up, or selection between for example to engage in ‘to participate in’ and to engage with ‘to establish a meaningful contact or connection with’. In these particular cases it might be possible to interpret the meaning of the expression correctly even if one has not encountered it before, i.e. to apprehend it passively even if it lies outside one’s active competence, but it is questionable how far most speakers ever stop to analyse idiomatic expressions such as to catch up on, to give (something) up, to leave off (doing something), on the one hand . . . on the other hand, to run (someone) to ground.

There is thus a very strong case for listing idiomatic expressions in dictionaries, so long as they are in sufficiently common use. They are often denoted technically by the broader term lexical item, as distinct from individual words or lexemes. However, not every lexical item that is listed in a dictionary automatically requires etymological investigation. We may feel that constructions such as to engage in and to engage with will normally be outside the scope of etymological research. However, some of the examples given above are less clear-cut. Understanding of the origin of the idiom on the one hand . . . on the other hand is helped by knowing that hand in earlier use had the senses ‘side of the body’ and more generally ‘side, direction’ (e.g. in an example from 1548 ‘on the other hand or side of the gate’6). The origin of to run (someone) to ground is understandable only when one realizes it originated in the specialist language of fox-hunting, referring to hounds running a fox to its burrow or earth. Many other idioms similarly rely on conventional metaphors which may or may not become opaque as a result of technological or cultural change, e.g. to run out of steam ‘to lose impetus or enthusiasm’ (which originated in the age of the steam engine) or to have shot one’s bolt ‘to have done all that one could do’ (which originated in the age of the crossbow), while others reflect otherwise obsolete or near-obscure senses of words, e.g. to cut a caper ‘to make a playful, skipping movement, to act ridiculously’ (showing cut in the sense ‘to perform or execute’ and caper ‘a frolicsome leap, especially in dancing’). Some originate in quotations, e.g. biblical quotations or paraphrase such as to turn the other cheek or to take someone’s name in vain, or quotations from Shakespeare such as the milk of human kindness or the world’s your oyster. (This last example becomes rather less opaque when the metaphor is heard in its original fuller context: The Merry Wives of Windsor II. ii. Falstaff: I will not lend thee a penny. Pistol: Why then, the world’s your oyster Which I with sword will open.) We will take up the difficult issues that such cases raise about the role of non-linguistic, encyclopedic knowledge in etymological research in chapters 8 and 9. Some idioms remain stubbornly resistant to all attempts to explain their origin, e.g. Bob’s your uncle ‘there you are’ (said in a situation where a task becomes easy to complete) or the full monty ‘everything which is necessary, appropriate, or possible, the works’.

Sometimes idioms arise from remodelling of earlier expressions. For instance, the rather opaque expression to have another thing coming (as in, If you think you can get away with that, you have another thing coming) becomes much more readily explicable when a little etymological research reveals that it is an alteration of earlier to have another think coming, in which think ‘action of thinking’ has been replaced by the commoner word thing (perhaps as a result of homophony in casual speech), even though the outcome is an idiom which is semantically much more opaque.

2.2 How new words arise

As well as looking at word forms and how they realize meaning, we can look at structure within the word, and in a book on etymology it makes most sense to do this primarily from the point of view of word origins, and thus to take a preliminary look at how new words enter a language.

2.2.1 Monomorphemic words and complex words

An important initial distinction is between monomorphemic words and complex words. As the name implies, monomorphemic words are composed of only a single morpheme or meaningful unit. Examples which we
encountered in chapter 1 include friar, sad, and deer: at least in modern English, these words are unanalysable units, and if we understand them it must either be because they are stored as meaningful units in our memory or because a given context in which they appear makes their meaning obvious. Other words are clearly analysable, such as happiness, steadiness, freshness, or closeness, although compare highness, which is analysable but not transparent, at least not in its use as an honorific title. It is important to note that it is not necessarily the case that these words are not also stored in our memory; but we can analyse all of them from their component parts (happy, steady, fresh, close, high, and the suffix -ness), and all except highness are semantically transparent. Throughout this section we will return often to the following questions:

(i) Do words of this type need to be included in an etymological dictionary?

(ii) Are words of this type interesting to etymologists?

We can immediately conclude that any monomorphemic words in a language will need to be included in any etymological dictionary which claims to be at all comprehensive, and that they will be of obvious interest to etymologists: from the point of view of the contemporary language they are stand-alone items which must have an origin and history which we will want to trace. A good case can also be made for including all affixes which are found in analysable words. (We will return to the etymologies of affixes in chapter 4.) The situation is much less clear-cut with words which are analysable, and we will need to look at a number of issues before we will be in any position to address this question.

2.2.2 Borrowed words

Words which have been borrowed from another language are typically monomorphemic, such as friar in chapter 1. However, some are analysable, usually because each of the elements in which they are composed have also been borrowed. For instance, English municipality is a borrowing from French municipalité, but it is analysable, because municipal has also been borrowed, and the ending -ity is familiar as the ending of a great many abstract nouns borrowed from French nouns in -ité (and/or Latin nouns in -itas) and has also become productive within English. Often it is difficult to determine whether complex words of this type show borrowing at all:

we will examine some of the issues concerned in sections 5.1 and 6.8. At a greater time depth, or where there is little data, borrowing generally becomes much more difficult to detect, and we will look at some of the implications of this in chapter 7.

Lexical borrowing is probably found to at least some extent in all languages, although the extent varies greatly (see chapter 5). We may fairly safely conclude that all words which have been borrowed will be of some interest to an etymologist, since we will want to find out how, when, and from which other language they have been borrowed. As we will see in chapters 5 and 6, these are very often difficult questions to answer, because of lack of evidence and/or difficulties of analysis. If we are even reasonably inquisitive about the ulterior histories of words, we will also want to delve further than this, and discover whether the word in the donor language is itself analysable and what its history is.

It may thus seem that all borrowed words will automatically need to be included in any etymological dictionary which attempts to be comprehensive. However, this presents some problems, both of a practical and of a theoretical nature. Fundamentally, words are borrowed, just as they are used, by individuals, not by 'languages', and we may find that very different selections of borrowed words belong to the vocabularies of particular social groups, geographical areas, etc., and even to the vocabularies of individuals within those groups, areas, etc.

Lexical borrowing is one of the many areas in which we can observe the open-ended nature of the lexicon of a language. Even if we restrict our focus to the usage of monolingual speakers, individuals have different interests or pursuits which will bring them into contact with different words from other languages. For example, very often people will have different enthusiasms for different cuisines, and accordingly they will have slightly different (active or passive) vocabularies of food terms. The Italian bread name focaccia has reasonable currency in contemporary British English, and also in many other varieties of English. The OED has an entry for this word as an English borrowing from Italian, with illustrative quotations dating back to 1881. However, the early quotations given in the OED present the word as an unusual item which authors feel the need to explain to their readers, and it is not until relatively recent years that we find examples reflecting more general currency of the word.

This particular example of a food term imported from another culture may seem an obvious symptom of modern cosmopolitanism and hence not
applicable to earlier historical periods, but in fact we find that imported items (foodstuffs, items of manufacture, etc.) are a very frequent source of new borrowings in almost all cultures and almost all historical periods. Inevitably, whenever we have a reasonably large body of historical data, we can ask, but not necessarily answer, the same sorts of questions about precisely whose vocabulary particular borrowed words may or may not have belonged to in a given place and time.  

Additionally, we should remember that mobility of individuals or groups between different speech communities is hardly a modern innovation, and much recent work in linguistics has highlighted just how typical (and indeed normal) bilingualism and multilingualism are in many parts of the world today and have probably been at all times in the past. We will look in chapter 6 at the rather vexed question of whether switches between languages by bilingual speakers actually show borrowing at all, and if not how great the connection between the processes is. However, as soon as we are dealing with a situation where people speak more than one language, it is fairly certain that there will be some interchange of lexis between the two languages, even if this is restricted to technical or specialist registers.

We can thus see that in any language a core of well-established borrowings is likely to be surrounded by a periphery of much less well-established ones. Wherever there is a language contact situation, any large sample of actual usage is likely to include nonce, one-off, borrowings which do not show more general adoption (although the same word may well occur as a nonce borrowing on multiple separate occasions).

The open-ended nature of the lexicon of any language becomes yet more apparent if we now consider new words which are formed within a language rather than borrowed from another language.

2.2.3 New formations: aspects of affixation and compounding

One very common method of forming new words is by affixation (or derivation). Both prefixes (which involve addition of material at the beginning of a base, e.g. un-, in-, pre-) and suffixes (which involve addition of material at the end of a base, e.g. -ness, -ment, -ly) are common in very many languages. We will look at both in detail in chapter 4. Much more rarely infixes are found, which interrupt a morphological base; in its inflectional morphology (rather than its derivational morphology) proto-Indo-European probably had an infix *-n- which formed present stems as part of its verbal system, reflected in for instance English stand.  

We sometimes also find circumsixes, which involve addition simultaneously of material at the beginning and the end of a base; by some analyses a circumsifix is shown by the ge- -r which is added to the stem of weak verbs in modern German to form the past participle, as e.g. gefragt 'asked', past participle of fragen 'to ask' (stem frag-), although again this belongs to inflectional rather than derivational morphology (unless we take the past participle to be an adjective formed on a verbal stem). 

In section 4.4.1 we will look at ablaut, the systematic employment of variation in a stem vowel to mark different morphological or derivational categories.

Another very common process is, as we have seen, compounding. One important thing that compounding and affixation have in common is that the resulting word is ‘bigger’ than the elements from which it is formed. The word form thus enacts the semantic relationship between a base word and a compound or derivative. When we encounter a new compound or derivative, we recognize that it contains a base word plus something else (an affix or another base word). This suggests to us that the new word will have a meaning related to that of the base word but modified in some way. This sort of relationship between word form and word meaning is termed iconic. (See further section 4.5.)

2.2.4 Productivity

If an affix is productive, i.e. capable of forming new words, it can sometimes generate an enormous number of new word forms. The process may be open-ended; this is particularly clearly illustrated by affixes which

---

7 For a detailed discussion of the general importance in etymological research of paying attention to how words can shift between specialist vocabularies and general usage see von Wartburg (1969) 107–14.

8 See Plag (2003) 101–4 for an argument that derivational inflexion is shown in modern English in expletive insertion of the sort shown by absobloomingately (see also section 2.1.1.4). On the distinction between derivation and inflexion see Plag (2003) 14–16.

9 Circumfixation should be distinguished from the simultaneous addition of both a prefix and a suffix in cases like decaffeinate < de- + caffein + -ate, where de- and -ate remain distinct affixes with distinct meaning and function. Such formations are normally called parasynthetic.

10 For a detailed analysis of morphological productivity see Bauer (2001); a useful account, with further references, is also given by Plag (2006). Productivity is a difficult and somewhat disputed term, and is not used in exactly the same way by all scholars.
can attach to names to form new lexical items, like *-ism in *Thatcherism, Stalinism, etc. New derivational formations may be formed at almost any time within the context of a particular utterance, and be understood within the context of that utterance. An influential study in this area is Baayen and Renouf (1996), in which the authors looked at frequencies of word forms with the affixes -ly, -ness, -ity, un-, and in- in the British newspaper *The Times* over a period between 1989 and 1993. They found very large numbers of forms which occurred only once in this corpus, and which were not recorded in any dictionaries. Their findings point strongly to very many of these formations being genuinely one-off nonce uses (examples include archdukelych, composerly, conductorly), which readers of the newspaper process effortlessly by means of their knowledge of the productive word-forming patterns of the language. These words are not stored in the reader’s memory, and yet they pose no problems for interpretation. Baayen and Renouf concentrated on words formed with derivational suffixes, but we can find just as great if not greater facility in the production of new compounds in English, which will be readily interpreted and understood by a hearer even if they are being encountered for the first time. (Of course, as noted in section 2.1.2, some scholars would anyway interpret at least some of these as showing phrases rather than compounds.)

Many words can be processed as they are encountered in context, drawing on the hearer’s or reader’s knowledge of the word-forming rules of the language. We can compare this to the way that any of an almost infinite number of different possible sentences can be interpreted (normally quite unconsciously) through the hearer/reader’s knowledge of the syntactic patterns of a language. Other words are stored in our memory, including some which are perfectly transparent and analysable. Some people will encounter and/or use some words regularly which some other people never encounter: Baayen and Renouf’s composerly, conductorly, and even archdukelych may be part of everyday discourse for some people. Many linguists invoke the concept of a mental lexicon, which will probably differ at least slightly for each individual speaker of a language.

If we take the view that an etymologist’s task is to account for the origin and development of the lexicon of a language, then this begins to appear

---

11 Additionally, they found that formations with the native, non-borrowed affixes -ly, -ness, and un- appeared to be much more frequent than would be suggested if one worked simply from the wordlists of dictionaries.

12 For an overview of this topic see Alitchison (2003).

an impossible endeavour if new words are continually arising in the speech or writing of different individual speakers and writers on a daily basis, and if different individuals will have different lexical items stored in their memories. A more useful framework for defining the main focus of an etymologist’s work is provided by the concepts of transparent and opaque (and also analysable and unanalysable) meanings and word forms which we have already encountered, and by the diachronic processes of institutionalization and lexicalization by which these commonly come about.

### 2.3 Lexicalization

A distinction is often made between nonce formations, institutionalized words, and lexicalized words. (More strictly, we should speak of lexical items here, so as to allow phrases to be included in the same framework.) Some scholars regard these as stages in a process which words may (but need not) undergo:

\[
\text{nonce formation} \rightarrow \text{institutionalization} \rightarrow \text{lexicalization}
\]

Nonce formations are ad hoc coinages by individuals in particular circumstances, the majority of which will never gain any wider currency, such as the words encountered in the Baayen and Renouf study which we looked at in the preceding section. Institutionalized words, while they remain (at least relatively) transparent, are used conventionally within a certain speech community in a given context or with a fairly specific meaning. Lexicalized words are opaque – in meaning, or composition, or both.

**lunchbox** is, compositionally, a transparent compound of lunch and box, and we are not surprised to find that it denotes a box for transporting one’s lunch. However, the definition in the *OED* suggests that it has some more conventional meaning characteristics than this:

A container designed to carry a packed lunch (or other meal). Formerly, any of various types and sizes of receptacle, sometimes also carrying crockery, etc., but now usually a small lidded box for food.

From the accompanying illustrative quotations in the *OED* we see that the modern use is most often specifically to denote such a box used for

---

13 See for example Bauer (1983) 45–50. For a thorough overview of this field see Brinton and Traugott (2005).
transporting lunch to a workplace or, especially, to school. This suggests that it is an institutionalized word for this item. If someone called the same thing a *foodbox or a *lunchcarrier we might understand from context what was meant, but it would strike us as not being the right word: in fact, it would be a nonce formation which we would interpret from the context in which it occurred, and we would soon conclude that it was intended as a synonym of the institutionalized word lunchbox. To take another example from the same semantic field, not many decades ago many British workers, particularly miners, carried their lunch in a metal container, usually called a snap-tin. Snap was a word for a light meal, and hence the compound was transparent, if institutionalized. However, today snap-tins (i.e. the physical objects) tend to be encountered only as collectables or museum pieces, and the word itself is encountered either as the name associated with these artefacts or in recollections of a bygone world. Internet discussions sometimes speculate on the meaning of snap in the compound, or feel the need to explain the word's origin. In fact it shows snap 'light or packed lunch', itself a metaphorical use of snap 'quick or sudden closing of the jaws or teeth in biting' (compare a bite to eat), which is in turn related to the verb snap. We could imagine an alternative scenario in which snap-tin was formed directly from the verb snap, perhaps because of its lid snapping shut when closing, and in which snap 'light or packed lunch' was so called because it was carried in a snap-tin; it is the historical record that shows us otherwise, rather than anything that we can intuit from the modern use of the word. Hence we see that for some speakers at least the term is not just institutionalized but lexicalized: they call this sort of box a snap-tin, but at least some of them are not sure why.

Lexicalization is an important process in any study of etymology, because it is key to explaining many word histories. In the case of snap-tin it is both the meaning and the composition of the word that have not become not just institutionalized but opaque: someone encountering the word lunchbox for the first time will have a good idea of what a lunchbox is simply from the composition of the word (even though they may miss some of the nuances of the institutionalized meaning), but someone encountering the word snap-tin for the first time is going to need to make careful use of information from the context of the wider utterance in order to work out what the word denotes, and will have little idea which out of numerous possible meanings snap shows in this word.

There are various different processes by which a word may become lexicalized. The most typical are:

1. Semantic change occurs, either in the lexicalized word or in one or more of its constituent elements (i.e. the words, affixes, etc. from which it is composed)
2. The word may become 'orphaned' as a result of one or more of its constituent elements becoming obsolete
3. Changes in word form (typically through the operation of sound change) may obscure the relationship between the word and its constituent elements

Often, more than one of these processes is found in a single word history, and it is sometimes hard to tell in what order they occurred. It is also often difficult to tell when a word became opaque, and a word may well remain transparent for some speakers when it is already opaque for others. Any change which results in the original morphological composition of a word becoming opaque is sometimes referred to as demorphemization or demorphologization (see e.g. Brinton and Traugott (2005) 52–4): for instance, in the case of handiwork which we encountered in section 1.3.2, the prefix ge- in the medial syllable has become opaque, as a result of loss of i- (< ge-) where it occurred word initially. (For further discussion of the prefix ge- see section 4.1.2.)

2.4 Examples of lexicalization

So far we have looked at lunchbox, a word which has an institutionalized meaning but is of transparent composition, and snap-tin, which is opaque for some speakers, but is also now a rather rare word. However, very many perfectly common words have shown a historical development from being analysable and transparent to being completely unanalysable and opaque. husband is a word with something of a 'disguised' history. As a modern English word it is unanalysable and indubitably monomorphemic, but this is not true at all points in its history. It occurs in its modern sense 'a man joined to a woman by marriage' from the thirteenth century. The word first appears, as late Old English húsbonda, in the eleventh century, in the sense 'the master of a house, the male head of a household'. It is a borrowing from Old Norse húsbóni (with assimilation to the class of weak masculine
nouns, hence the ending -ə in the nominative case in Old English). However, the composition of húsbonda would have been transparent to speakers of Old English, since the first element hús is identical in form and meaning to its Old English cognate hús ‘house’, and the second element bônđi ‘peasant owning his own house and land, freeholder, franklin, yeoman’ was also borrowed into late Old English as bonda (i.e. again with assimilation to the class of weak nouns). Indeed, as with many borrowed compounds, it could alternatively be argued that húsbonda was formed in Old English from hús and bonda on the model of Old Norse húsþondi (see section 5.1 for discussion of this topic, and also 5.2 for terminological complications to do with the term ‘Old Norse’). In the Middle English period the vowel in the first syllable of the English word was shortened as part of a regular process of shortening before consonant clusters. Consequently it did not participate in the Great Vowel Shift affecting long vowels, as house did, with the result that the first element of the word became opaque, since hús- /huz/ (later /huz/ or /huːz/) showed no obvious relation to house /haus/. (We will return to the Great Vowel Shift in section 7.2.3.) Old English bonda is continued by Middle English and early modern English bonde, bond, but the word is now obsolete. husband has thus become opaque as a result of:

- semantic specialization
- formal change in its first syllable (and different formal change in the parent word house)
- obsolescence of the word which forms its second element

As is typical in such cases, it would be very difficult to identify exactly when the word ceased to be transparent. If we consider that a language is something spoken by large numbers of individuals, we can see that it will be impossible ever to pin down a precise moment when change occurred, because the relevant changes in word form and word meaning will not have occurred for all speakers at the same time. In fact, the evidence of spelling forms and recorded meanings in the OED suggests considerable overlap both between different meanings and between different forms in the history of this word, just as we find in a great many other cases as well. Additionally, if we are trying to assess whether people in the past perceived a word as a transparent compound, we will always be engaging in guesswork to some degree: we can show that in such and such a period the language contained relevant word forms, so that someone so minded could make the connection between simplex word and compound word, but we cannot demonstrate that this actually happened. Thus we may in some cases be able to show when a word ceased to be analysable, but we cannot show when it ceased to be analysed. (It can be difficult to gauge whether a word is perceived as a transparent compound even by contemporary speakers.)

To take another example, English lord was also originally a compound, even though in modern English it is both monomorphemic and monosyllabic. It is recorded in Old English most commonly in the form hlæðfērd, but also once in the form hlæðfweard. It has a range of meanings in Old English, including ‘master’, ‘prince’, ‘chief’, ‘sovereign’, ‘feudal superior’, and even ‘husband’, but probably its original meaning was ‘the male head of a household’. Although poorly attested, hlæðweard is almost certainly the earlier form of the word, showing a compound of hlæð (modern English loaf) and weard ‘keeper’ (modern English ward); the original meaning was thus metaphorical, referring to the role of the head of a household as owner and provider of the food eaten by his servants and dependants. In the more usual Old English form hlæðfērd with reduced second syllable the connection with weard is already obscured, and very possibly no connection with hlæð was felt either. Certainly, all formal connection with loaf is lost in the reduced monosyllabic form lord which becomes the usual form from the middle of the Middle English period. lady (Old English hlæðdige) probably shows a similar origin, < hlæð + an otherwise unrecorded word with the meaning ‘kneader’ ultimately related to dough. (In this instance hlæð in the Old English word form shows the sound change known as i-mutation: see section 7.2.4.)

In each of these cases changes in word form have played a major part in making the etymologies and early meanings of the words opaque, i.e. demorphologization has occurred. In other cases change in meaning is much more important than change in word form. The word handsome is formed from hand and the suffix -some. This suffix seldom produces new words in modern English: it has become unproductive and now only occurs in occasional analogous nonce formations. The words in which it survives are a rather complex set of lexicalized words in which the suffix shows a number of different relationships with the base word, e.g. quarrelsome, bothersome, loathsome, fearsome, wholesome, cumbrous. However, in all of these cases it remains clear that e.g. quarrelsome has some connection with quarrel or quarrelling, and bothersome with bother or bothering, even if a particular speaker is unfamiliar with the lexicalized meanings ‘given to or characterized by quarrelling’, ‘annoying, causing bother’, etc. In some
other cases the parent word has simply become obsolete, as in the case of winsome (from Old English wynn ‘joy’); viewed synchronically, it has become unanalysable and so a unique morph (more commonly called a cranberry morph, for reasons we will see in section 2.6). In the case of handsome the situation is rather different. The first element is hand, and this is still very clear from the written form of the word. There is often no ld/ in the spoken form, but careful listening shows that the same applies to handshake, handsaw, hands-off, hands-on, handstand, and other words with a similar sequence of sounds, as pronouncing dictionaries will confirm, and yet in all of these cases the relationship with hand remains perfectly obvious. The crucial difference in the case of handsome is the development in meaning that the word has shown. When first found in the fifteenth century the word meant ‘easy to handle or manipulate, or to wield, deal with, or use in any way’, and in the early sixteenth century also ‘handy, ready at hand, convenient, suitable’ (we may compare the semantic history of handy already investigated in chapter 1). But these senses are now obsolete in most varieties of English, and the word has passed via the senses ‘apt’, ‘proper’, ‘fitting’ to the core modern senses ‘especially of a man) good-looking’, ‘of a number, sum of money, etc.) substantial’. In consequence all semantic connection with hand has been lost, and the word has become opaque.

penknife presents an interesting case of a word which is perhaps rather less far down the route of lexicalization. It obviously and transparently denotes a type of knife. However, to the vast majority of modern speakers, it does not have any obvious or transparent connection with pens. The Oxford Dictionary of English (revised edition, 2005), a dictionary which takes a synchronic (i.e. non-historical) approach based on a corpus of contemporary usage, boldly defines penknife as ‘a small knife with a blade which folds into the handle’. It also offers no etymology for the word, and in my view this could conceivably leave some readers confused about its origin; they might guess wrongly at some connection with pen ‘small enclosure for animals’ (reasoning that penknives have some sort of basic out-of-doors function), or they might assume that this kind of folding pocket knife was invented by someone with the surname Pen or Penn. Or perhaps they will alight on the right pen, but with the wrong reasoning, assuming that a penknife is a knife which is taken to resemble a pen when folded away. This is perhaps a little unlikely, but most people will probably need to engage in a little lateral historical thinking to arrive at the right answer. It is much more likely that in the ordinary course of events they will give the matter no thought at all, and regard penknife as the specific but inherently uninformative name of a type of knife. The historically based definition in the OED (third edition, entry published 2005) informs anyone about the history of the word penknife immediately: ‘Originally: a small knife for use in making and mending quill pens (now rare). Now usu.: a pocket knife with one or more blades (and occas. other tools) designed to fold back into the handle when not in use.’ And to avoid any lingering confusion, a brief formal etymology is provided, identifying that the word is indeed a compound of pen ‘writing implement’ and knife; hence pen has in formal terms an objective relation to knife, denoting the thing which the knife is (or rather was) used to sharpen. In this case it is the changing use of the denotatum, i.e. technological change in the non-linguistic world, which has been the driving force leading to lexicalization.

A final example will introduce some further themes which we will explore more fully later in this book. The word acorn is clearly monomorphic and unanalysable in modern English. Furthermore it has a satisfying meaning relationship with an easily identified and very tangible entity in the real world. If someone asks us what the word acorn means (or more likely, what an acorn is) we can point to an acorn and say ‘it means one of these’. (Although a botanist may note that different types of oak tree in fact have different types of acorns.) However, etymologically the word acorn is almost certainly related ultimately to the word acre, the modern reflex of Old English acre ‘field’. It probably originally had the meaning ‘fruit of the unclosed land, natural produce of the forest’, although by the date of its earliest recorded appearance in English (in the form aceren) its sense has become restricted to ‘acorn’, the fruit of the oak tree, to which the authoritative Dictionary of Old English adds ‘perhaps other fruit of similar form, mast’ (that is to say, the fruit of woodland trees, such as acorns, beech mast, etc.). The meaning development, and the relationship between acorn and acre, become clearer when we look at some of acorn’s cognates in other Germanic languages: Dutch aker ‘acorn’, Old Norse akarn ‘acorn’, Old High German ackeran ‘oak or beech mast’, Gothic akaran ‘fruit’. We have no real way of knowing for certain whether the Anglo-Saxons connected the word with acre, but the restricted meaning, and the lack of any metalinguistic comments to the contrary, would suggest quite strongly that they did not. In modern English both the word’s meaning and its form disguise the etymological connection with acre, and etymological investigation is required to establish the connection and to trace how the two words subsequently diverged. Interestingly, the word has been subject to various
folk-etymological alterations during its history in English, indicating a desire on the part of language users to establish iconic relationships with other words in the language. (See further section 7.4.5, and also 4.5 on iconicity.) In the seventeenth century we find the form *oke-corn, in which the word has been remodelled after oke, a variant of oak, and corn. Thus the word's form has been altered in such a way as to make transparent a perceived basic meaning 'corn (or fruit) of the oak', which certainly reflects what an acorn is, but this does not coincide with the word's historical composition. The modern form acorn (rather than *akern) results from this same folk-etymological association with corn.

2.5 Apparent reversals of the process

Very occasionally the interaction between the written language and the spoken language may lead to apparent reversal of the lexicalization process. This typically happens in languages which have a standard and long-settled written form. The written language may therefore not reflect changes in word form which have occurred since. Thus breakfast, blackguard, or boatswain all reflect their composition transparently in the written form, but not in the spoken form (/bɾeɪkfast, /blækgɑrd, /boʊsən/), although since blackguard and boatswain are both now relatively rare words 'spelling pronunciations' are sometimes heard for each of these, hence /blækgɑrd/ or /boʊtswən/ (but /boʊtswən/ would never occur as the spoken realization of the adapted spelling bosun). Such spelling pronunciations can sometimes completely oust an older pronunciation which shows demorphologization, hence /wɛstkɑʊt/ rather than /wɛskɪt/ is now usual for waistcoat, and /fɪʃhɛd/ is becoming more common than /fɔːrd/ for forehead. We will look in section 7.4 at various other processes such as folk etymology which run counter to lexicalization, since they lead to an increase in compositionality and analysability, and which are therefore sometimes described as showing anti-lexicalization.14

2.6 Cranberry morphs

If compounds and derivatives are common in a language (as they certainly are in English), this can lead to a certain degree of tolerance of words which have the appearance of being compounds or derivatives but in which one of the elements is not analysable. The first element of the word cranberry is totally opaque to a speaker of modern English who does not know something about the history of the word, and morphologists often refer to unanalysable morphemes of this kind as cranberry morphs (or alternatively, and less colourfully, as unique morphs).15 In fact, the word cranberry has been opaque for all of its history in English. It shows a seventeenth-century North American English borrowing from another Germanic language, probably Low German, in which the word ultimately shows a cognate of the bird name crane and a cognate of berry; compare the forms Low German kranheere, High German Kranheere. In English, the second element of the word has been remodelled after, or perhaps assimilated to, the English cognate berry. As a result the word belongs to a family of words denoting types of (relatively) soft fruit, which also includes such transparent formations as blackberry and blueberry which both have fairly clear reference to the characteristic appearance of the fruit, although both are clearly institutionalized names. (Someone might hypothetically perceive blueberries as being more black than blue in colour, but that person could not then reasonably expect to be understood if she began to refer to blueberries as blackberries without making it very clear that she was making a deliberate departure from conventional linguistic usage.) Various shrubs of the genus Symphoricarpos (most of them originally native to North America) are normally called snowberry in English. Many of these have white berries, and this might seem the obvious reason for the name, but some others have red berries. The name may simply have been transferred from the white-berried type to the red-berried type, and indeed the white-berried type do appear to have been the first to be given this name. However, most snowberries, regardless of colour, bear their berries in winter, and this might suggest a quite different motivation for the name, or alternatively explain how the name could easily be transferred from the white-berried to the red-berried type, if reanalysed as referring to the season when the plants bear their berries. The reason for the strawberry being so called is far from obvious; it is normally considered by etymologists that it shows the word straw 'stem(s) or stalk(s) of various cereal plants', but various explanations have been suggested to account for this, such as the appearance of the plant's runners, or the appearance of the small seeds on the surface of the fruit, or perhaps the name reflects the cultivation of strawberries

---


on beds of straw to keep the berries off the ground. *Raspberry* is almost certainly a compound of the earlier word *rasp* denoting a raspberry, but without a knowledge of linguistic history we may just as well think that *rasp* is a clipping (or shortening; see section 4.4.3) of *raspberry*; compare some fruiters' use of *straws* for *strawberries*. Thus we see that within this group of words we have a cline of different degrees of analysability: *blackberry* and *blueberry* are obvious descriptive names; *snowberry* may be a less certain case; *strawberry* may be analysable if we stop to think about it, but is hardly likely to be apprehended as a descriptive name in everyday use; *raspberry* may be a longer alternative name for *rasp*, but in synchronic terms the two words are merely synonyms and *rasp* is of no aid in explaining *raspberry* since we do not know the origin of *rasp*; *cranberry*, so far as its existence in English is concerned, is evidently a type of berry, but has a first element with no connections elsewhere in the language, unless we happen to know its further etymology in Low German and work backwards from that to the English cognate *crane*, but that is purely extralinguistic knowledge. If we consider the different types of fruit which these various plants have, it also becomes clear that the concept denoted by *berry* in these formations is not a very precise one; we will return to this point when we consider prototype semantics in section 8.2. Nonetheless, the group of words ending in -*berry* has acquired new members through folk etymology: *naseberry* denoting the sapodilla (a type of fruit which grows on a tree) in fact shows a borrowing from either Spanish *náspera* or Portuguese *nêispera*, with the ending remodelled by folk-etymological association with words ending in -*berry*. (On this etymology compare sections 7.4.5 and 8.8.1. For some further *berry* names see section 9.7. A further interesting example to pursue is *gooseberry*.)

2.7 Which words need etymologies?

We have seen that the lexicon of any language will be extended by speakers in an ad hoc way, as new words are formed by productive word-forming processes such as derivation or compounding. These will normally be understood very easily by other users of the language from their transparent composition and from clues in the context of the utterance which help to explain the meaning. Only a tiny percentage of such introductions are likely to be adopted more widely. If we are working on a dead language or an ear-

lier historical stage of a living language which has a relatively small corpus of surviving material, then we may decide nonetheless to include all of the surviving words in any etymological dictionary or corpus of etymologies, on the not unreasonable assumption that not enough evidence survives to enable us to see which formations are completely trivial and transparent and which are not, and it is therefore much better to be safe than to be sorry. If we are working on a contemporary language, we will certainly not have this luxury. Since the lexicon is almost infinitely extendible, it will be impossible for us to compile a comprehensive list of all of its words, let alone etymologize all of them. But this poses a problem for etymologists: as we will see in subsequent chapters, investigating almost any word history involves either implicitly or explicitly drawing parallels with other word histories, and we will not want to run the risk of neglecting words which may provide crucial information in explaining another etymology.

A useful framework for deciding which words to concentrate our energies on is provided by the concepts of transparency, opacity, and analysability, and by the insights provided by observing the diachronic processes of institutionalization and lexicalization. We might decide that our ideal etymological coverage of a language will include:

- any monomorphemic words (although we may need to reconsider this in the case of languages where variation of the stem vowel is a productive method of realizing derivational relationships: see section 4.4.1)
- any word containing a cranberry morph
- any word which has a form which is not explicable by the productive word-formation processes of the language
- any word which is formally analysable but semantically opaque, e.g. *handsome*, *handy*, or for some speakers *penknife*; also idioms such as *to cut a caper*

The last category is particularly difficult to define, since what is opaque for one speaker may not be for another. In each of these categories, our etymological investigations will in many cases show that the current status of a word results from earlier lexicalization, as e.g. *lord*, *lady*, *acorn*, *strawberry*.

We may also decide to add:

- all remaining words with a non-predictable, institutionalized meaning
- all phrases and constructions with institutionalized meanings not readily predictable from the meanings of the words of which they consist
3

Are words coherent entities?

3.1 Variety in form and meaning: poke ‘bag, sack’ 62
3.2 Do we know precisely when a word’s history begins? 68
3.3 Homonymy and polysemy 74
3.4 How polysemy–homonymy relations can change 76
3.5 Merger (or near-merger) in form and meaning 79
3.6 Splits in word form 83
3.7 A case of merger followed by a split 86
3.8 Homonymic clash 88

In this chapter we will look more closely at the variation in form and meaning shown by individual words. Variation may be found within a single time period and locality, and between different historical periods (diachronic variation) or different geographical areas (diatopic variation). We will take up the crucial issue of how words change with time, and we will examine critically whether we can always take it for granted that a word has continuity as a coherent unit from one historical period to another. We will look at the importance of tracing the process by which a word has developed. In comparison with this, the actual point of origin may be a relatively trivial matter, although we will also look at some cases where it is very difficult to pin down exactly when a particular word originated. We will examine cases where two originally separate words have merged, and conversely cases where one individual word has split into two or more separate words. In doing so, we will gain a better understanding of the data which forms the basis of etymological research, and will be better prepared for a more detailed investigation of the major issues in etymological research.