



This is not a squirrel (see p. 44).  
Detail from *Allegorical "Millefleurs" Tapestry with Animals*,  
c. 1530–45, Bruges, Belgium. *The Minneapolis Institute of Arts*,  
Gift of Mrs. C. J. Martin in Memory of Charles Jairus Martin.

# WORD ORIGINS

*... and How We Know Them*

Etymology for Everyone

ANATOLY LIBERMAN

KEISER COLLEGE - TALLAHASSEE  
850 - 906 - 9494

OXFORD  
UNIVERSITY PRESS

2005

A browser of an English etymological dictionary may conclude that some time around the sixteenth century a combined Dutch-German invasion followed the Viking raids and the Norman Conquest. The number of Middle Dutch words, homonymous with their cognates in Northern German (or Low German, as it is called in linguistic works), that entered English and became fully domesticated in it is astounding, and they are not limited to seafaring, warfare, or trade. Most of the frequentative verbs discussed in Chapter 9 are of Dutch or Low German origin. *Brackish, drawl, drill, groove, loiter, snip, snap*, and hundreds of others have the same source. The dictionary in which they are featured is 660 pages thick.<sup>29</sup>

No modern European language has received so many words from so many languages as has English.<sup>30</sup> Whether this openness has always been a blessing is a matter of opinion. Foreigners groan under the burden of English vocabulary. Native speakers, who, as time goes on, read less and less of their classical literature, understand it worse and worse. Both language and literature develop by canonizing their lower (popular) forms and rejecting some of the achievements of past epochs. Fewer and fewer people remember the difference between *timorous* and *temerarious*, but at the moment, everyone knows the difference between *nerd* and *geek* and between *awful* and *awesome*, and that is the way it has always been. Most words branded in Samuel Johnson's 1755 dictionary as low are now respectable, whereas Shakespeare must be read with a sizable glossary. Some borrowings had their day and disappeared, others stayed and gladden both a discriminating user and an etymologist who know the difference between *doughty, bold, stalwart* (from English), *stout* (from Anglo-Norman, originally Germanic, like German *stolz* [proud]), and *brave, valiant, valorous, courageous, and intrepid* (from French), and are not afraid to add the English suffix *-less* to both Engl. *fear-* and French *daunt-*. Mastering a language, even one's own, especially such a rich language as English, is a gallant deed.

## Chapter Thirteen

*in which the plot does not thicken, or*

### A Retrospect: The Methods of Etymology

Language changes, but we take no notice.—Internal reconstruction.—Cognates, congeners, and other family business.—On galleys and galleries.—Chuck Taylor endorses Converse brand tennis shoes.—Cognates versus borrowings.—The first summing up.—Say no to look-alikes.—The more, the better.—If possible, stay at home.—A waif arouses pity.—From things to words.—Good wine needs no bush.

My story is approaching the culmination, and the time has come to throw a retrospective glance at the strivings and achievements of etymology, the better to appreciate the revolution in historical linguistics that will be discussed in the next chapter. Etymology finds its justification in the belief that words are, or, at a certain stage in the development of language, were, not arbitrary but meaningful combinations of sounds. Every decipherment presupposes that the code can be broken; in this respect, an etymologist is like a decoder.

Words change both their phonetic shape and meaning (see especially Chapter 2). This is not a trivial statement. We understand the oldest people around us and our great-grandchildren, and the ease of communication emphasizes the stability of language. Some words appear and disappear in our lifetime, stress can shift from the second syllable to the first, and usage does not remain the same from decade to decade, but those are details not comparable with social upheavals, revolutions in the style of clothes, and the collapse of age-old taboos. The paradox of language is that it changes fast and radically, without our noticing it. *Barn* (from *bere* + *ærn*) and *daisy* (from

*dæges ēage*) have lost half of their sounds. Several centuries ago, *stone* was pronounced with the vowel of *store* and before that with the vowel of *spa*. *Book* and *fight* had the vowels of modern *Bork* and *feet*, respectively. Most of the oldest words of English are now monosyllabic (*see, speak, cat, dog, head, mouth*, and so on; see p. 74) and often remain such in declension and conjugation (*saw, speaks, dogs, mouthed*), in contrast to what they were in Chaucer's days, let alone in the days of Hengist and Horsa, the semi-legendary leaders of the fifth-century Germanic invasion of Britain. Evidently, to be successful, etymologists should try to uncover the oldest recorded form of the words they are researching.

Human memory is short, and our historical intuition, when it comes to words, is unreliable. We learn with surprise that the *Oxford English Dictionary* has no citation for *fake* (called slang) before 1812 (can the word be so recent?) and that *floozy* was known in 1911 (could it be around so long?). The naive idea that one can discover the origin of a word by looking at it attentively and thinking hard yielded to the demand for studying Old and Middle English, Old French, and so on. The first professional etymologists in Western Europe (and they appeared about four hundred years ago) were partly aware of the situation, but their knowledge of the earlier stages of the modern languages was limited; proficiency in Latin, Greek, and Hebrew could not make up for that gap in their education. With few exceptions, the grammars, dictionaries, and editions in our libraries do not antedate the middle of the nineteenth century. The amateurs who, as late as 1900, filled the pages of popular journals with their conjectures on word origins, had no idea that they should have used their time reading rather than writing.

The earliest attested English words return us to the beginning of literacy in Britain. Some of them were coined after 450, but most were brought by the invaders to their new home from the continent. They are millennia away from any primordial utterances and monosyllabic grunts that allegedly mark the rise of human speech. However, we must be grateful for what we have. Without the evidence of Old Engl. *hūswīf* (literally "housewife") and *heahfore*, etymologists would have been hard put to reconstruct the history of *hussy* and

*heifer*. Not everybody is so fortunate. For untold centuries, all culture was oral, and in many cases, historical linguists have no texts to work with and depend entirely on so-called internal reconstruction. For example, they note words like *to husband* and *husbandry* and conclude that "the male spouse" could not have been the first meaning of the noun *husband*. Every language has such significant accretions from the past, but they seldom go far back. Or we compare several modern languages and try to guess which forms are more archaic. (Thus we can compare Engl. *do* and German *tun*, and decide that at one time English infinitives ended in *-n* but later lost it.)

Terms like *Germanic* and *Romance* presuppose the existence of language groups. The members of a group are related, that is, they go back to the same ancestor and share certain features inherited from that ancestor and absent elsewhere. The parent of the Romance languages is Latin (the substrates are taken for granted), and both French and Spanish etymologists must be fine Latinists. The parent of Germanic (the Scandinavian subgroup, English, Dutch, Afrikaans, German, Yiddish, and a few dead languages like Gothic) has not been recorded. No texts exist in it, and here our position is less advantageous, but the situation with Romance is an exception: all the other protolanguages are the product of reconstruction. A historian of Germanic words compares forms from the languages belonging to the group and, considering how long Germanic speakers were the neighbors of the Celts and Romans, needs more than a smattering of Celtic and Romance linguistics. In practice, all-encompassing erudition is rare, but the ideal remains.

A term that has frequently occurred in the pages of this book is *cognate*, a noun or an adjective (for example, *flatter* is cognate with, or a cognate of, German *flattern*). In the same sense, *allied*, *akin*, and *related to* have been used. The noun *cognate* has a synonym *congener*. We spot some cognates without any difficulty, for example, Engl. *house*, German *Haus*, Dutch *huis*, Swedish *hus*. Others are less obvious: Engl. *tooth*, *four*, *love* and *go* versus German *Zahn*, *vier*, *lieben*, and *gehen*. Cognates are offspring of a protoword (reconstructed but not attested, whence the warning asterisk), which may stay ossified in one language and change its sounds and meaning in

another. Icelandic *hús* (*ú* designates long *ú*, as in Engl. *who*) has the same pronunciation today that it had 2,000 years ago. Of the two words—*tooth* and *Zahn*—the German one has preserved its shape better than its English cognate, but it, too, has lost a consonant (the original form of *Zahn* was *zand*).

Selecting cognates is an indispensable first step of every etymology, but a string of related forms may not solve the question about the word's "nonconventional" meaning. Our aim is to learn how the combination of sounds *h-ū-s* came to mean "dwelling." By discovering *Haus* and *huis*, we do not come any closer to the solution, though we realize that the diphthongs in English, German, and Dutch developed by later phonetic processes: in medieval texts, the word appeared as *hūs*. We are in better shape with Engl. *flatter* and German *flattern* (to flutter), because thanks to the German cognate (assuming that it is indeed a cognate of the English verb and of *flutter*), the shift of meaning becomes clear: from "fluttering" around the person whose good graces are our objective to insincere praise. Note the parentheses in the previous sentence: it is helpful that *flatter* and *flattern*, unlike *house* and *Haus*, mean different things, but this benefit has a shady side: we are now no longer certain that the words we paired are cognates. Thus, we either face an uninspiring set of words nearly or wholly identical in form and meaning (*house* ~ *Haus* ~ *huis* ~ *hús*) or clusters like Engl. *flatter* / *flutter* ~ German *flattern*, whose members are not necessarily related to one another. Every attempt to find an etymology depends on the selection of cognates. Is *clover* really akin to *cleave* (to stick to), and *chide* to *kīd* (wedge)? I think they are, but not everybody will agree, because the connection between *clover* ~ *cleave* and *chide* ~ *kīd* is not self-evident.

The passage on *galley* (p. 139) ended with the statement that *galley* is not related to *gallery*. To make such a statement, one has to investigate the history of both words. *Galley* was originally a Greek animal name. When the letter *G* in the *Oxford English Dictionary* was going to print, nothing worth repeating was known about the origin of *gallery*. Since that time, Romance scholars have made a few suggestions, and *The Oxford Dictionary of English Etymology* was able to state the following: "Perh[aps] alteration of *galilea* GALILEE . . ."

*Galilee* (a porch or chapel at the entrance of a church) derives from the name of a province of Palestine, "perh[aps] used in allusion to it as being an outlying portion of the Holy Land; first recorded of Durham cathedral and taken up thence by antiquarian writers of [the nineteenth century]." The origin of *gallery* is still unknown. A porch is not a gallery, and in the wanderings from Medieval Latin (*galeria*) to Italian (*galleria*), northward to French, and to English, a good deal of information has been lost. Earlier researchers did connect *galley* and *gallery*, but they based their conjecture on a wrong etymology of *galley* (which they traced to Greek *kālon* [wood; lumber; fleet]).<sup>1</sup> With that etymology discredited, we can say that *gallery*, despite its obscurity, is not a cognate of *galley*.

It is clear why isolated words are the hardest to etymologize, though isolation is a relative concept. *Chide* (if *kīdel* does not belong with it) is only English, *clover* (if unrelated to *cleave* [to stick to]) is limited to a few Germanic languages, *house* (if all the non-Germanic cognates proposed for it are wrong) occurs only in Germanic, unlike *father* or *eight*, with connections from Norway to India. It is such isolated words that may be fragments of a substrate. As pointed out, the one precious cognate we need may not have been recorded or the obscure word we are studying was coined in a way incomprehensible to us.

For instance, the development of *ragged* to mean 'exhausted, edgy' and used to refer to males shows that the original allusion to menstruation, *on the rag*, has been lost. The slang *chucks* for 'high topped tennis shoes' seems entirely arbitrary without the information that Converse brand tennis shoes carry the endorsement of Chuck Taylor. The verb root of *wanker* 'loser' is puzzling without the information that *wank* is an imitation of the sound of the buzzer on Truth or Consequences.<sup>2</sup>

Hundreds, if not thousands, of words similar to *chucks* and *wank* must have been in circulation for centuries. Our chance of guessing their origin is slim. *Ship*, in the politely subdued jargon of modern researchers, is a word of doubtful etymology, which means that the hypotheses on its origin advanced so far are not fully convincing, though the comparison with Latin *scīpio* (staff, pole) is not bad (a

vessel we call "ship" may have been a hollowed log or a dugout). But perhaps its etymon (the form from which it is derived) meant "basket" or "can." Words for such containers are numerous. Even we, with our miniscule command of Old English vocabulary, know about two dozen; at least as many may have been lost. Yet the problem remains: to discover the origin of a word, we need cognates with comparable but nonidentical meanings.

Compounds and words with prefixes and suffixes form a special group. *Bridal* goes back to a sum of two nouns: *bride* and *ale*. *Nimble* is *nim-b-le*, and *balance* is traceable to *\*bilancia*. Nothing else can be said about them. The next step would be a search for the origin of *bride*, *ale*, *nim*, *-b-*, *-le*, *bi-*, and *lancia*, but the results of that search are of no consequence for understanding *bridal*, *nimble*, and *balance*. In dealing with such words, etymology merges with word formation. Transparent words like *undo*, *shipment*, and *statesman* do not interest etymologists, who step in only when questions arise, for example, about what is *orna-* in *ornament*, *-couth* in *uncouth*, *cran-* in *cranberry*, and *straw-* in *strawberry*, and why we say *spokesman* rather than *\*speaksman*.<sup>3</sup>

In theory, cognates (descendants of the same parent) are easy to distinguish from borrowings (guests from another language). But let us look at two examples. A thousand years ago, English and German had the word *hūs* (house). Today, English and German have the word *nylon*. English and German *nylon* are not cognates. Engl. *nylon* was invented in 1938 by the DuPont Chemical Co., a coinage that makes one think of the textile (compare *rayon*) and perhaps vaguely suggests the fabric's novel character (compare *new*). Both the product and its name became popular in many countries. Is it possible that a special type of dwelling called *hūs* originated in some one Germanic language and spread to its neighbors, as *nylon* did in the twentieth century?

The earliest Germanic word for "house" seems to have been *razn* (recorded in Gothic). Its cognates have been mentioned above in connection with Engl. *barn*, *saltern*, and *ransack*. The Germanic *hūs* was, in all likelihood, different from the *razn* (compare the differences between Modern Engl. *house*, *building*, and *edifice*). In Gothic,

a language recorded 16 centuries ago, *hūs* occurs only in the compound *gud-hūs* (godhouse), that is, "temple." *Hūs* is not a maritime word, but its origin is no less obscure than that of *ship*. Since a prehistoric counterpart of the DuPont Chemical Co. is hard to imagine, we assume that *hūs* is not a borrowing in any of the older Germanic languages but a reflex of a Proto-Germanic word. Although this assumption is justified, it cannot be proved. In the thirty-seventh century, someone who will write a book like the present one may suggest that *nylon* is a Proto-European noun consisting of the negation *n-*, the root of the Greek word *hýle* (forest)<sup>4</sup> (with *h* dropped), and a suffix of probably substrate origin, the whole meaning either "containing no fibers" or "not to be worn in a wooded area." We have seen that Engl. *crab* and Russian *korob* (basket) may be either native words in their languages or borrowings. Engl. *garden* is a cognate of Slavic *gorod* (town), but borrowing (from Germanic into Slavic or from Slavic into Germanic) is not inconceivable. "The great problem of comparative philology is to distinguish between those resemblances which are the result of common parentage and those which are the result of influence, or what is called 'borrowing'."<sup>5</sup> Every student of historical linguistics comes to the same conclusion.

An etymologist deals with probabilities. As long as we have the support of documents, we are historians. *Heifer* undoubtedly developed from *heahfore* because the Old English form *heahfore* and its later reflexes (continuations) have been recorded, but this form is opaque. Why did it mean "a one-year-old cow that has not calved"? Both *ea* and *o* could be short or long. To produce an etymology, we will try to choose the most promising variant of four (*hēahfore*, *heahfore*, *hēahfōre*, *heahfōre*), though opinions about what is promising differ.

*Heifer* is a notoriously hard case, but probability is the foundation of most etymologies. *Snark* is certainly, not probably, a blend of *snake* and *shark*, because Lewis Carroll explained his coinage. By contrast, Swift did not bother to tell us why he called his little people *Lilliputs*, and we are not better off with that name than with *heifer*. The name of the person who introduced *slender* into English is irrelevant. The probability of a blend (*slight* + *tender*) is rather high, but

it is still only a probability. Although unanimity is rare among etymologists, the degree of their success is impressive. The origin of thousands of words has been discovered and codified in excellent dictionaries. The mechanisms of phonetic and semantic change (to be discussed in the next chapters) and the role of the ludic element (language at play) are today understood so much better than they were even two centuries ago that the science of etymology can be proud of its achievements.

From the foregoing exposition a few principles of etymological analysis have emerged. It may be useful to summarize them and list them in one place.

- Etymology does not depend on look-alikes. Engl. *house* and German *Haus* are similar (nearly identical) and related, whereas *galley* (from Greek) and *gallery* (a reshaped Hebrew place name?) are similar and unrelated. On the other hand, Engl. *tooth* and German *Zahn* belong together, though today all their sounds are different. Their relatedness or the lack thereof can be established only by comparing the oldest extant or reconstructed forms of each word. Folk etymology suggests ties based on chance resemblances. It will explain *gossip* as *go sip* and invent a plausible yarn about how *sirloin* originated in the phrase *Sir Loin*. Indulging in amateurish fantasies should be discouraged (which does not exclude the possibility that someone without any training in linguistics may know a story or a local custom of real value to an etymologist: see the explanation of *chucks* and *wanker*, above).
- An etymology that can “decode” several words is, in principle (note the hedging), preferable to the one that offers a separate explanation for each word of what seems to be a set. For example, if the choice is between an etymology of *Lilliputian* that fits only this word and an etymology that sheds light on both *Lilliputian* and *Laputa*, it is advisable to accept the second one. However, a hypothesis that purports to explain dozens or even hundreds of words is usually suspect. The immutable law—the

broader the volume, the more narrow the content—is valid for all formulations. Hence the danger of hearing onomatopoeias and ideophones everywhere, overstating the role of sound symbolism, detecting blends in all obscure words, and in reducing the entire vocabulary of a language to a few roots (the last point, like the mechanisms of change, will be discussed later).

- It is often unclear whether a word is native or borrowed. In such cases, the probability factor plays an especially important role. *Crab* may be of Germanic descent, but the existence of numerous similar words elsewhere makes the idea of borrowing more appealing (not proven but only more appealing; our evidence is inconclusive by definition: words are not characters in a Conan Doyle or an Agatha Christie story and are not in a hurry to confess even on the last page). A bad etymology is not better (in fact, it is much worse) than no etymology at all (“origin unknown,” a resigned acquiescence in inevitable ignorance, as Jespersen put it in a Micawberian way), but given a high probability that a word has an ascertainable origin in its language, caution is needed in suggesting a foreign source, be it a neighboring language, the language of ancient colonizers, or an unidentifiable substrate.
- When a word occurs in several languages and the question arises where it originated, its home should be sought in the language in which it has ties with other words. This is why *zigzag* seems to have been coined in German and *hackney* in English.
- Every word was coined by a resourceful individual or borrowed as a result of language contact in a certain place at a certain time. It has an etymon, a sound complex endowed with meaning. Some words are short-lived, others become a permanent part of the vocabulary. Newcomers may oust their synonyms that have existed for centuries. The staying power of words increases if they form ties with other words. *Bob* (an insect) merged with the name *Bob*, *gun* from *Gunilda* (assuming that



this derivation is right) pretended to be a clipped form of Old French *mangonne*, so that upstarts began to look like old-timers. This should not be interpreted in the sense that *bob* and *gun* have two etymologies each, but it means that the survival of a word may depend on the soil from which it springs up. Our inability to choose among several equally reasonable solutions should not be used as a plea for the ability of a word to have multiple etymologies.

- The knowledge of things around us cannot be derived from words (or names, as Plato called them), but the sidelight from etymology occasionally illuminates the past. If *ship* is really cognate with Latin *scīpio* (staff, pole), this fact confirms our notion of the most primitive sailing vessels. However, in research, the process starts at the opposite end: to arrive at a plausible etymology of *ship*, we must have an idea about primitive ship building. Etymology is not about the word's "true meaning," because any meaning acceptable to a given community is "true." Its goal is to break through the conventional nature of the linguistic sign. When success crowns this endeavor, *cuckoo* emerges as an onomatopoeia, *balance* as "two weighing scales," and *lord* as "the guardian of bread."
- As a general rule, a good etymology is simple (only finding it is hard). Name givers use the material close at hand: a husband is a homeowner, a lady is a bread kneader, a galley is a weasel, and a cloak is a bell-shaped (clock-shaped) garment. Etymologies presupposing many complicated moves need not be wrong but usually are.

It appears that we already know a good deal about an etymologist's work. Yet the most rewarding part of the story lies ahead.