Anatoly LIBERMAN, University of Minnesota

What Can We Expect from a New Dictionary of English Etymology?

Abstract

The field work for a new dictionary of English etymology is drawing to an end. In the course of this work, the format of the dictionary has been clarified and a vast bibliography collected. The prospective dictionary will have the following distinctive features. 1. It will offer a critical survey of scholarship on every word included. 2. It will appear in an unusual order: a. Words without established cognates outside English; b. Words presumably Germanic but without established cognates elsewhere; c. Words with broad connections in Indo-European; d. borrowings from the non-Germanic languages. 3. Detailed discussion of isolated words (group a) will reveal the processes of language creativity that have remained stable from the most ancient times.

Keywords: English, dictionary, etymology

Although there is an overabundance of English dictionaries with the adjective *etymological* in their titles, English post-Skeat lexicography cannot boast of outstanding achievements. 17th and 18th-century authors tried to guess the origin of words and sometimes succeeded. The discovery of the comparative method, editions of Old and Middle English monuments, and researches into living dialects supplied etymology with a solid base. Although inspiration and good luck still play a role in etymological work, our guesswork has become more sophisticated. However, English etymological dictionaries do not keep abreast of publications in their field, and, on the whole, their level is that of Skeat[4] (1910) and the *OED*, unfortunately, without their brilliance. Apart from antedating, only more recent words occupy the attention of etymological editors on the staff of "thick dictionaries", and it is a characteristic fact that Burchfield decided not to revise the etymologies going back to Murray, Bradley, and Craigie. Judging by *The Oxford Dictionary of English Etymology (ODEE)*, Onions (1966) did not have too much to say on this subject either. He codified the material from the *OED*, added a few new derivations (without explaining why he had done so), and made his entries a model of bland, noncommittal, and uninspiringly safe scholarship.

In 1987 I began working on a new dictionary of English etymology. Its main goal is to offer etymologies worthy of the present state of English, Germanic, and Indo-European historical linguistics. Outside the area of English studies, most authors of etymological dictionaries take it for granted that an entry should contain references to the hypotheses of their predecessors. Walde, Feist, Vasmer, Frisk, von Wartburg, Jan de Vries, Corominas – to mention just a few – would have refused to write their works in the dogmatic, apodictic style adopted by their English colleagues. As long as etymology was tantamount to listing conjectures, all of which allegedly had equal value, lexicographers promoted stocktaking; cf. such relatively late dictionaries as Johnson-Todd (1827) and Richardson (1858). But Wedgwood (1859–65) was much more economical, and Skeat referred to an article or a book only when he could not help it, Brugmann's *Grundriß* being the only notable exception. Murray introduced the policy of avoiding unreliable etymologies, and if none of the current derivations convinced him, he would say, "Origin unknown". Predictably, he was unable to live up to his principles, for most etymologies in all languages are "uncertain" or "debatable", but, however many

references Murray's team and Skeat might give, they signified a forced retreat from the program considered ideal. Later lexicographers followed the same principles.

It is the duty of a good etymological dictionary not only to offer the least controversial solution but also to present the case, sift all the conjectures, dismiss fantasies, and pave the way for further research. The prospective dictionary will fulfill precisely this task. At the moment, its database contains about 13,000 titles of chapters in books, articles, and reviews in more than a dozen languages, with information on approximately as many English words. Among other things, the bibliographical component of the dictionary will bring to light the mole-like activities of the etymologists working for our best dictionaries. For example, *boy* is a word of dubious antecedents. For a long time it was believed to be of Germanic origin, but Dobson (1940, 1942) traced it to French. This etymology, which is most probably wrong, has been accepted at Oxford and elsewhere. An analytic entry will make it clear that Middle English *boie* derived from Anglo-Norman **abouié* is not the "truth" but one scholar's hypothesis.

In any language with a long recorded history, some words have easily recognizable cognates, while others appeared as if from nowhere. Some words of English belong to the Indo-European stock, some have no ties outside Germanic, and some are isolated (it is not even clear whether they are native). Very many English words were borrowed, especially from French and Latin. The prospective dictionary has primarily the needs of English philology in view. It is interesting to investigate the numeral *eight* (< **ok'tou*) as a possible dual and connect it with the name for the breadth of a palm (four fingers twice) or for the harrow (an implement with eight teeth), but an English etymological dictionary is not needed for this purpose: Gothic *ahtau*, Sanskrit *aštáu*, etc., lend themselves much better to such speculations. Likewise, it is useful to explain that *lethargy* is a borrowing from Old French, which in turn continues late Latin *lēthargia* < Greek *lēthargíā* from *lēthargos* 'forgetful' < **lēth* (as in *lethal*, Greek *lēthā*'oblivion'), variant of **lath-* in *lanthánein* 'escape notice', probably related to Latin *latēre* 'be hid', but most of this information belongs in dictionaries of Latin and Greek. Even the best experts in the history of English words seldom have original ideas on the protoform of *eight* and on the relations between Greek *lanthánein* and Latin *latēre*. But when it comes to words like *boy, girl, cub, slum, snob, big, pig,* and so forth, it will not do to dismiss them with the verdict "origin unknown". If an English etymological dictionary has nothing to say about them, who will take up this subject? We are informed by Onions that the connection between *pig* and Dutch/Low German *big* 'pig' "cannot be made out". Actually, it can, but this is not the main point here. It is more important to note that the literature on the etymology of *pig* is extensive, and, unless dictionaries of English and Dutch leave the convoluted prehistory of *eight* and *lethargy* to Indo-Europeanists and turn their attention to *pig* and *big*, these words will forever remain in limbo.

In suggesting reasonable self-restraint, I am following the tradition of Romance linguistics, from Diez onward. The well-known problems notwithstanding, most nonslangy, nonborrowed words of French, Italian, Spanish, and Portuguese can be shown to go back to Latin. However, the etymological dictionaries of these languages ususaly take the reader only as far as the asterisked Romance root. The origin of the Latin word must be looked up elsewhere. Unlike Romance, Germanic has no recorded parent and some discussion of the most ancient words in it is unavoidable, but here, too, a measure of self-restraint would be welcome. As stated above, investigating the exact form of the Indo-European etymon of the word *eight* need not be English etymologists' first priority. Nor should they attempt to commit

themselves to a particular brand of laryngeal theory. Given the enormousness of "local" problems, a historian of English vocabulary should preferably stick to his last. This also holds for borrowings. Etymological lexicography is poor in theoretical works, and its practice is mainly based on precedent. Thus, it is customary to give detailed accounts of the remote antecedents of English words borrowed from the Classical and Romance languages, but supply other borrowings with a mere hint of derivation. My project has not advanced far enough for me to be able to formulate hard and fast rules concerning loanwords, but it seems that, as a matter of general principle, an English etymological dictionary should concentrate on the history of the borrowing in English, rather than on the history of its source, and leave French etymologies (to give the most important example) not to paid consultants, but to compilers of French etymological dictionaries. I would make an exception only for the words borrowed by English from other Germanic languages, for within Germanic, national borders are often blurred.

Very little has been written about an optimal size of an etymological dictionary and of its corpus. An ideal dictionary should of course list all the words of the language and explain their origin. Derivatives, especially those with nonproductive affixes, are also worthy of consideration: cf. the verbs *startle* and *fizzle* and the nouns *girdle* and *handle*. But in practical work, one should sometimes be guided by common sense, rather than by lofty principles. The origin of an English word is so easy to find in popular dictionaries that an editor active at the end of the 20th century can afford the luxury of being selective. An etymological dictionary is a highly sophisticated reference tool. Thousands of people want to know where the words of their language came from, but in most cases they expect a simple, straightforward answer, and *Webster's Collegiate Dictionary* or *The Concise Oxford Dictionary* will give them all the information they need, while an etymological dictionary can concentrate on issues that it alone can solve. This is particularly true of numerous exotic words. On a randomly chosen page of *The ODEE* (496), we find *jihad, jinn, jinricksha, jirga(h), jiu-jitsu* (which refers us to *jujitsu*), *jocko,* and *joey* 'young kangaroo'. The entries devoted to these words contain reference to the source language and state when they appear in English. This seems to be a waste of space. Nor is it immediately obvious that *jigsaw, jim-jam,* and *jobation* are valuable additions to an etymological dictionary. Finally, I have doubts that *jink, jinx, jitter,* and even *job* 'piece of work' enhance the value of *The ODEE,* for we are only told that the origin of these four words is unknown (about *jitter* nothing at all is said: it is only glossed). Discussion would have been most welcome, but Onions offers none. If we also remove *Job,* who, not unexpectedly, turns out to be "a patriarch of the O.T. taken as a type of destitution and of patience", *Jock* (a variant of *Jack*), and the reference words *jill* (see *gill*), *jimmy* (see *jemmy*), *jod* (see *yod*), and *jodel* (see *yodel*), we will be left with *jiminy, jingle, jingo, job* 'pierce', *jobbernowl, jockey,* the dialectal word *jockteleg* (included for unclear reasons), *jocose, jocund, jodhpurs,* and *joey* 'fourpenny piece', that is, with 11 words out of 31. Of these rescued entries only the one on *jingo* is of considerable length, while the others are between three and seven half-lines long. Publishers are happy when they can promise their prospective buyers a truly representative dictionary, but they may not realize how trivial the product they admire is. I believe that, given the extremely high level of English lexicography, the next etymological dictionary of English etymology should contain as *few* words as possible, each supplied with detailed discussion.

For the purposes of the prospective dictionary the vocabulary of Modern English has been divided into several major groups: 1) words without established cognates outside English, 2) words with ties in one or more other Germanic languages, but without established cognates

outside Germanic, 3) words with connections (in and) outside Germanic, 4) unquestionable borrowings. The publication of the dictionary, which will proceed in fascicles, will start with group 1 and finish with group 4. It is to be expected that the new dictionary, in addition to presenting a critical history of scholarship, will lift numerous English words of "unknown (uncertain) origin" from their obscurity. Judging by what has been done, the obscurity of at least some of them has been greatly exaggerated. Not a single etymology has been or will be written with the view to offering an original solution, but knowing practically everything scholars and amateurs have said about an English word and all its cognates in the course of three centuries and a half often makes it possible to formulate new hypotheses. They may not be startlingly new (an enriched synthesis is perhaps the best name for my production), but the entries certainly go beyond a survey of other people's views. Besides that, the reader will come away from the dictionary armed with a near-complete bibliography of the etymologies in question. It is hard to exhaust the material I set out to investigate, but in the absence of any cumulative bibliography of English etymology, the references offered in the dictionary will be sufficient for a good start.

And a last point. The dictionary will be able to present the process of language creativity in a new light. One example will suffice. The origin of the English words *big, bag,* and *bug* 'insect' poses serious problems. *Bag* is a doublet of *pack* (Old Icelandic also had *baggi* and *pakki*). *Big* at one time meant 'strong, stout'. *Bug* is even more difficult, for it was first recorded in the 16th century. Etymological dictionaries hedge when it comes to their deriva- tion. But let us suppose that *big, bag,* and *bug* were Old English and Old Icelandic words and found their way into (Walde-)Pokorny's dictionary. They would undoubtedly have been analyzed as representing different grades of ablaut of the same root, with the meaning 'prone to swell; swollen'. Ablaut never lost its productivity in Germanic, so even though *big, bag, bug* surfaced late, they can also be etymologized along these lines. And when we discover that English dialectal *bug* means 'big', *bog* means 'boastful', while Norwegian *bugge* is glossed 'big man', our hypothesis gains in credibility. Nor should the extremely common variation *b-, p-* bother us. It is customary to set up Indo-European doublets like **bimb, *bhimb*. But **bimb* and **bhimb* are a linguistic fiction whose existence depends on the attested forms *pimp ~ bimp!* We can safely connect not only English *pig* and Dutch *big* but also Dutch *big* and English *big*; after all, pigs have been fattened for millennia.

Bug is especially troublesome also because it seems to be a variant of Old English *budda*. Once again we deal with a case that would not have given (Walde-)Pokorny any trouble: *-g* and *-d* would have been classified as extensions appended to the root **bu-*. However, extensions are paper gimmicks, for pure roots did not function as independent words: we witness an *alternation* of the *bug- ~ bud-* type, rather than the formation *bu+g, bu+d*. In light of the variants *budda ~ bug* we can explain the connection between Old English *padda* and Low German *pogge* 'frog, toad, pad(dock)'. Middle English *budde* 'bud' belongs here too. Buds and frogs are endowed with a capacity to swell and need not be separated from (big) pigs, bags and bugs.

Observations to this effect have been made many times, but, curiously enough, the forms that would seem natural to an Indo-European scholar baffle students of English, and dictionaries keep telling us that *big, bag, pack, bug, bud, pad(dock)*, and *pogge* are of unascertained origin. This profession of ignorance rests on the belief that only old languages were truly alive, whereas in a modern language the bridge from *b-* to *p-* and from *-d* to *-g* cannot be crossed. Indo-European is allowed to have variation, but modern languages are sustained on

the strict diet of regular correspondences and sound laws. Owing to the fact that the prospective dictionary will not shy away from obscure words, it will reveal the ties long since known to dialectologists and semioticians but persistently ignored by the authors of etymological dictionaries. The emphasis laid on "languages at play" (there even exists a term "ludic forms") should not be understood as the betrayal of Neogrammarian algebra, but, like so many other things, this algebra is meant to be a servant, not the master of our endeavors. Entries will be long. It is more reasonable to explain *bag – pack – big – bug – bud* together than to split this group into six or seven partly overlapping articles. Cross-references and a word index will make the whole material easily accessible to the user. The sample entry below will give an idea of the format of the dictionary (a book containing about 6,000 words). The text will be printed in two columns single-spaced (the format of the *ODEE)*. There will be a full bibliography at the end of each fascicle, but in the present paper the references occurring in the sample will not be deciphered. *Eena* is not a typical etymology, for it lacks the otherwise very strongly represented philological component, but it is the shortest written so far.

EENA (1855)
This is an ancient Celtic numeral ('one') still current in English, especially in Yorkshire, in counting sheep (along with similar words for 'two,' 'three,' 'four,' and 'five'). These pseudo-numerals were brought to New England and used as tally marks in trading with Indians. Now they are preserved only in children's games.

Eena is dismissed by OED as a nonsense word. In AMG (250), the jingle, "Eena, meena, mina, mo, Catch a nigger by the toe, If he hollers, let him go, Eena, meena, mina, mo" is called comparatively recent, without further specifications. An old exchange of opinions on 'the ancient British numerals', known better among students of folklore than among etymologists, reveals the prehistory of *eena*. Here are the first numerals used in scoring sheep in Yorkshire Dales and transcribed with so-called Glossic signs: *yaan, taih'n, tedhuru, (m)edhuru, pi(m)p,* i.e., [jain], [taiðn], [teðð͡rð], [(m)eðð͡rð], [pi(m)p] (Ellis [1870:117; 1871: XIX]). I. Taylor's (1877:338) list of "ancient numerals which were formerly in use in the northwestern corner" of English is similar: *eina, peina, para, pattera, pith,* etc. In his opinion, "these numerals are a relic of a language of the British kingdom of Strathclyde or Cumbria, which stretched northwards to Dumbarton, and whose southern boundary ran a few miles to the north of the place from whence these numerals have been obtained." He adds that, according to a local tradition, "the numerals were brought to Craven by drovers from Scotland. This tradition in no way implies that the numerals are Gaelic, but may be sufficiently explained by the fact that a great part of the Cumbrian kingdom lay to the north of the modern Scottish border." Ellis traced the Yorkshire numerals to Celtic, namely, to "the Welsh branch, dreadfully disfigured in passing from mouth to mouth as mere nonsense". But Bradley (1877) wondered how Cymric numerals "could have become so familiarly known in Yorkshire" and believed "that they had descended traditionally from the time when a Cymric dialect was spoken in that district"; he looked upon them as ancient British, rather than Welsh. The same in Bradley (1879). In this belief he was supported by Taylor. All the materials were published in the same volume of *The Athenæum.* According to the editorial note (p. 43), *The Athenæum* received "a great many more communications on the subject" than the magazine could print. See also Westwood (1877), Ellwood (1877), Powell (1877), and Trumbull (1877). Later authors (e.g., Beddoe and Rowe [1907:42]) repeated Bradley's conclusions. For a contemporary account see Potter (1949–50b).

The numerals recorded by Taylor, Ellis, and Bradley are sometimes mere gibberish, with English words replacing the original form (cf. *yahn* = [ja:n] 'one,' the local pronunciation of *one*) and rhyming words invented by informants. A similar string of numerals was in use among some North American Indians, e.g., *een, teen, tother, fither, pimp*, with the variants *eeny, teeny, tuthery, fethery, fip*. A list of Wawena numerals from Maine first appeared in print in Brunovicus (1868:180), with reference to a communication by R.K. Sewall dated Winter 1867. Kohl (1869:91) suggested in passing that these numerals "bear a resemblance to the Icelandic" (which they do not). Trumbull (1877) corrected Kohl's mistake and pointed out that these scores were "to be regarded rather as *tally-marks* or *counters* than as true cardinal or ordinal numbers. They were employed in counting off by fives, tens or twenties. Traces of some such systems may be found in many school-boy rhymes for 'counting out'." (pp. 14–15). In his opinion, the supposed Indian numerals were "brought to New England by English colonists and used by them in dealing with the Indians in counting fish, beaver skins, and other articles of traffic. When the memory of their origin was lost, the Anglo-Americans believed them to be Indian numerals, and the Indians, probably, believed them to be good English." Other variants of the rhyme in question abound (Newell [1853:194–203], Bolton [1888:103–08, nos. 568–646], Macritchie [1915, esp. 282], Witty (1927:44–45), Cassidy (1958:23–24), and Abrahams and Rankin [1980: nos. 119–411]).

According to Potter (1949–50a), the second line of the rhyme goes back to French-Canadian *cache ton poing derrière ton dos* 'hide your fist behind your back'. Misunderstood by anglophone children, it allegedly turned, under the influence of their parents' conversations when the Fugitive Slave Law of 1850 was being debated, into *catch a nigger by the toe*. He does not exclude the role of an Indian or a half-breed intermediary and points out that in the earliest variants *(eena, meena, mona, mite, basca, lora, hora, bite, hugga, bucca, bau; eggs, butter, cheese, bread, stick, stock, stone dead – O–U–T)* there is no mention of Negroes. The French-Canadian hypothesis is interesting but hardly necessary (O&O [1951:156–57]). Even less credible is Potter's imaginative reconstruction of "an ancient magic rime-charm used in Druid times to choose the human victims to be ferried across the Menai Strait to the Isle of Mona to meet a horrible fate under the Golden Bough of the sacred mistletoe amid the holy oaks" (340).

Not only sheep are scored in the way described above, and not only in Yorkshire; see the examples in OED and in O&O (1983:12–13), but the first word *eena / eenee / eny / ina* is the stable element everywhere, even in German. Levin (1995:422–23) surveys the opening line within the framework of his theory of displaced numerals. *Eena / eeny* looks like a numeral that made it all the way from the relics of the Cymrian dialect "in the Yorkshire dales" to the pidginized English of Indian trade posts in North America. If we do not follow Potter all the way to the Golden Bough, *meeny* will appear as a corruption of *teen*, while *mina* and *mo* seem to be nonsense words alliterating with *meeny* and leading up to the pair *mo / toe*. (Liberman [1994b:175–78].)

It is often asked to what extent the projected dictionary will be "popular". It will be neither more nor less popular than its numerous scholarly counterparts. Some etymologies can be presented in the form of so-called word histories. *Eena* is a prime example of this situation. Other cases are more difficult. Lay readers do not know the difference between Old Franconian and Old French, do not distinguish Old Saxon, West Saxon, and the dialects of modern Saxony, and are unaware of the periodization in the history of even their own language. Etymology cannot be discussed without reference to the grades of ablaut, breaking,

and so forth. Outstanding scholars pretend, for commercial purposes, that they are writing for Everyman, a situation which would be unthinkable in the sciences. Quite predictably, they fail. Here is Elmar Seebold's entry *Mus* in Kluge-Seebold (1995:576), with the bibliography left out. "**Mus** *n.* (<9. Jh.). Mhd. *muos,* ahd. *muos* 'Essen, Speise, Mus', as. *mōs* 'Speise, Essen' aus wg. **mōsa- n.* 'Zukost', auch in ae. *mōs,* afr. *mōs* 'Speise, Essen'. Offenbar eine Vriddhi-Bildung zu dem auch als *s*-Stamm auftretenden Wort (g.) **mati-/ez* 'Speise' (s. *Messer).* Das Grundwort hat die Tendenz, 'Fleisch' zu bedeuten (vgl. ne. *meat),* die Ableitung steht für 'Gemüse, Brei u. ä.', regional auch 'Obstbrei'." All Germans probably remember how *"Mus"* tastes, but how many of them will feel comfortable in the forest of the abbreviations Seebold uses, understand the meaning of the macron, or have the faintest idea of the s-stem and Vriddhi? At this stage, we need a solid dictionary of English etymology. Later it can be turned into many books, booklets, and even looseleaf calendars for the popular reader.

References

Dobson, E.J. (1940). The Etymology and Meaning of *boy. Medium Ævum* 10, 121–154.

Dobson, E.J. (1943). Middle English and Middle Dutch *boye, Medium Ævum* 12, 71-76.

Johnson, Samuel (1755). *A Dictionary of the English Language.* Printed by W. Strahan, London. Revised by H.J. Todd. 2nd ed. London: Longman, etc., 1827.

Kluge, Friedrich (1995). *Etymologisches Wörterbuch der Deutschen Sprache.* 23rd ed. by Elmar Seebold. Walter de Gruyter, Berlin, New York.

OED = The Oxford English Dictionary. Clarendon Press, Oxford, UK. 1884–1928. 2nd ed., 1992.

Onions, C.T. (ed.), with the assistance of G.W.S. Friedrichsen and R.W. Burchfield (1966). *The Oxford Dictionary of English Etymology.* Clarendon Press, Oxford, UK.

Richardson, Charles (1858). *A New Dictionary of the English Language, Combining Explanation with Etymology.* Bell and Daldy, London.

Skeat, Walter W. (1910). *An Etymological Dictionary of the English Language.* 4th ed. Clarendon Press, Oxford, UK.

Wedgwood, Hensleigh (1859–65). *A Dictionary of English Etymology...*Trübner & Co., London, 2nd ed., 1872; 3rd ed., 1878; 4th ed., 1888.