ANATOLY LIBERMAN University of Minnesota-Twin Cities, Minneapolis (MN) aliber@umn.edu

THE FUTURE OF ETYMOLOGICAL DICTIONARIES

Keywords: etymology, dictionary, bibliography, reconstruction, borrowings

Abstract

Now that printed books are being replaced by online materials, it is especially important to agree on the format of the etymological dictionary of the future. It seems expedient to discontinue the publication of dictionaries that contain minimal or no new information, for the public already has more than enough of them. The profession needs exhaustive (ideally annotated) bibliographies of everything ever published on the origin of every word in the language under study. Of great use can be thematic etymological dictionaries, such as dictionaries of presumably native words in a given language, of borrowings, of slang, of regional words, etc. Only the languages that have never been the object of sustained etymological research require general dictionaries of the type once produced by Skeat, Kluge, and their peers.

In the not too distant future, all traditional reference tools will be superseded by digital resources. Printed books will disappear long before we agree on the origin of Go.¹ boka 'letter' or Russ. kniga 'book'. Already now the literate world spends most of its time online and words are searched for rather than looked up. Old etymological dictionaries have yielded to countless pages available from the Internet. In preparing for the days when libraries will become computoriums, we should give some thought to the future of publications in our field and to the state of the art on the threshold of the epoch whose arrival will coincide with festivals of book burning along the lines familiar from November 5 in Great Britain, with dictionaries

¹ The following abbreviations are used below: Go. – Gothic, Engl. – English, OE – Old English, Russ. – Russian.

replacing the effigies of Guy Fawkes. One good result of those conflagrations will be that the length of our dictionaries will no longer matter.

The present essay is, as announced in its title, devoted to the future of etymological dictionaries, or, more properly, to their format. The public wants to know where words come from. Professional linguists do too, but, unlike inquisitive amateurs, specialists have to decide how far they can and should go and how much they can say. The oldest authors of the modern era (beginning roughly with the seventeenth century) assumed the existence of the protosource of all languages. Some found it in Hebrew, others preferred Dutch. Such monomaniacs, to use Ernest Weekley's term, are still with us. Dictionaries by deluded authors "prove" that all words of their favourite language can be traced to Hebrew, Arabic, Russian, or Irish Gaelic. Reading their works is sometimes amusing, sometimes sad and even frightening.

Our old predecessors were fortunate in that they had a clear view of their point of departure. In contrast to them, we attempt to *reconstruct* that remotest point. We know where we want to find ourselves, but our Promised Land is, to paraphrase slightly the title of Ehm Welk's novel, *das Land, das ferne leuchtet*. For some it is Proto-Indo-European; for others, Nostratic. Naturally, all reach their place of destination. Whether that place has anything to do with reality is another matter. In what follows we should not lose sight of two most important questions: "How much of that dreamland's territory has to be covered in an etymological dictionary?" and "To what extent are we ready to commit ourselves?" I will begin with the second of them.

Some words have minimal chronological depth. Such is nearly all slang. Occasionally researchers succeed in discovering the area in which a slang word was coined and the time of its first appearance, but more often they hit the wall. Obviously, *dude* did not exist in Proto-Indo-European or even Old English. Other words of undiscovered origin are older, though not necessarily very old. They are not exotic or slangy, or vulgar, and share common ground with *dude* only in that they too came in from the cold. About all of them dictionaries say "of unknown origin". Here are a few English words whose early history is said to be shrouded in obscurity. The numbers in parentheses refer to the centuries of their first attestation. A much longer list can be found in Liberman (1992).

Adz(e) (OE), awning (17), akimbo (15), balderdash (16), bamboozle (18), basket (13), binge (19), blurb (20), boy (OE), brag (13), breeze (OE), coax (17), dandy (17), dig (OE), dodge (16), doldrums (19), dwarf (OE), ever (OE), fad (19), fidget (16), fog (16), garbage (16), girl (13), goblin (14), guess (17), heifer (OE), hint (17), hire (OE), hobo (19), inkling (14), jog (14), jump (17), lad (13), loom (OE), mongrel (15), nod (14), oat (OE), pond (13), pony (17), qualm (16), quip (16), rabble (16), scoundrel (16), shallow (15), sham (17), sliver (14), sprawl (OE), stooge (20), strawberry (OE), toad (OE), trash (16), wench (13), yeoman (14).

When we approach such words, we cannot know whether they are all native (a few of them are certainly not). Their arbitrarily chosen distinctive feature ("of unknown origin") ignores their homeland. It only indicates that their beginnings are lost. The question naturally suggests itself: "Is there anything to say about them?" Indeed, there is.

The formula "of unknown origin" conceals a variety of situations. Some words are totally obscure, that is, no one has a clue to how they arose (this is a relatively rare case). Much more often dictionary makers prefer to sit on the fence because no agreement on the word's derivation has been reached. Two, three, or more intelligent guesses may compete, with none of them carrying conviction. However, knowing them is important, for even a dubious hypothesis may contain a grain of truth or show a later researcher the way to a better solution. Even the most conscientious lexicographer is often unaware of a work that sheds light, however dim, on the problem at hand. Of the words cited above I have dealt with *adz(e)*, *awning*, *boy*, *ever*, *girl*, *hire*, *hobo*, *lad*, *loom*, *oat*, *scoundrel*, *strawberry*, *toad*, and *yeoman*. I did not come up with definitive answers but succeeded in disproving some unmaintainable conjectures (including a few that enjoyed nearly universal support) and offering tentative solutions. In at least three cases (*hire*, *loom*, and *yeoman*) I probably even solved the riddle. Meillet's witty dictum that all good etymologies have already been found and the new ones are bad is discouraging and wrong.

Years ago, I realized that the post-Skeat, post-OED dictionaries of English etymology do not reflect the state of the art and depend on outdated, non-representative databases. The literature on most words is hard to collect, and lexicographers, if they are not the peers of Feist (VWGS), Vasmer (REW), von Wartburg (FEW), Hofmann (WH), Mayrhofer (KEWA), and Trubachev (ESSI), who devoted years or even all their professional lives to the production of etymological dictionaries, have limited or no knowledge of the monographs, articles, and notes pertaining to the words they describe. After more than two decades of work on my own dictionary (see Liberman 2008; a mere introduction) I have ample confirmation of that view. Excellent suggestions about the history of English words turned up in books, popular journals, reviews, and articles in about twenty languages (see Liberman 2010; since that time this bibliography has grown very considerably). I am convinced that etymological dictionaries of the future should abandon the phrase "of unknown origin" and offer instead exhaustive critical surveys of what has been written on the subject. The absence of a word in a dictionary would mean that the author either has not dealt with it or has nothing to say.

In the future, condensed dictionaries that include most of the vocabulary of a given language will make sense only if there is no or almost no national tradition behind them; that is, not for English, German, French, Spanish, and their likes. The general public will be satisfied with reprints of what we already have, while specialists and other serious students will need topical dictionaries. One of the first "installments" may be "Words of presumably unknown origin". Its publication should be prepared for by a database like the one I put together for English. The editor will have read all the works included (for comparison: Kennedy [1927] did not have to read the articles and books featured in his bibliography; even those who write for *Year's work* have no time to think deeply of what ends up on their desks and in their computers).

Early in my etymological career, I gave a talk on my plans to the Philological Society (Oxford). Professor Terry F. Hoad noted that it would be good if I published

a summary of the literature I had amassed. I explained to him that so-called "interim etymologies" exist but that doing what he had proposed for the entire database would take thousands of hours and pages. I have not changed my opinion, but I now see a practical way of following his advice. Synopses are possible, but they should be made available step by step. A volume like Feist's *VWGS* (3rd ed., 710 pp.) might be sufficient for the most basic English or German words "of unknown origin", though Feist often had to dismiss a hypothesis without discussion, and in the work I envision this approach is inadmissible. It will not do to say that a certain etymology is wrong; each conjecture has to be discussed before being buried. I'll briefly return to this point in the conclusion.

Words borrowed in the course of roughly the last millennium, that is, since the beginning of writing in post-Roman Europe, pose a grave problem. It is unrealistic to expect that a specialist in English historical linguistics (I cite English as an example; the same can be said about other languages) should feel equally at home in Frisian, German, and Scandinavian (Yiddish constitutes a special difficulty, more formidable for a German than for an English scholar), but for the sake of argument let us assume that English etymologists do indeed range freely over all the Germanic languages. Even if such people exist, they will never feel equally comfortable in Romance. Etymological lexicography of the future is thinkable only as a venture with participants from several fields (possibly from several countries).

I have studied the history of the English words *apricot, bar, baron, barricade, beggar, bigot, brave, brothel, ghetto, marquis, petty,* and a half-dozen others of the same type. In doing it, I followed every footnote, probably missed very little, and opened every existing dictionary, but I can have no independent opinion about the dialectal forms of French, Italian, Spanish, and Portuguese. My knowledge of Romance lacks depth, and I have not developed an almost instinctive feeling for what is right and wrong in French, the instinct that guides me through English, though I can read linguistic literature in Romance languages and Medieval Latin. If such is the situation in Romance, what can I say about Celtic? An accident of birth allows me to deal with Slavic; most of my colleagues in the West cannot boast of this advantage. To repeat: some "installments" will have to be team work.

We have turned the words *diversity, interdisciplinary*, and *global* into meaningless clichés, mere tokens of conformity to the ever-changing political climate. I wonder what organization will agree to fund a small diverse, interdisciplinary, global, and highly sustainable group of specialists (hardly more then ten full-time coworkers), whose goal will be to produce an etymological database and a series of durable etymological dictionaries for the main languages of Europe. In the absence of such a utopia, dedicated students of English (again citing English only as a case in point) should probably try to take care of borrowings in this language themselves. Since I have already made use of the title of one German novel, I will risk summarizing a modern etymologist's experience by referring to another, this time "borrowed" from Hans Fallada: "*Jeder stirbt für sich allein*."

At this juncture, I can address the first question formulated at the beginning of this paper. So far, we have dealt with words that etymologists hoped to trace

to other recorded words. One could, for example, ask: "Does Engl. adz(e) go back to Go. aqizi 'ax' or any of its cognates?" As long as researchers try to answer such questions, they stay in the real world. But etymology has always tried to break into the realm of phantoms. Those phantoms are roots, and we need to know what place roots should occupy in an etymological dictionary and especially in an etymological dictionary of the future.

The root as a concept of historical linguistics is an ambiguous or perhaps double-edged concept: it is understood as both the common part of a group of forms (for instance, *cry*- is the root of *cry*, *crier*, and *decry*) and the begetter of related words, whose relatedness is deduced from the existence of this still undiscovered root (such are the asterisked entities in WP and *IEW*). I discussed this ambiguity in Liberman (2003) and need not go into detail again, for I am only interested in whether abstracting the root is a desired or even the ultimate goal of an entry in an etymological dictionary.

A search for reconstructed roots in Western languages was at one time inspired by the existence of triconsonantal roots in Hebrew. The emergence of what we now call scientific philology, contemporaneous with the discovery of regular sound correspondences (sound laws) and the works by Jacob Grimm, changed the procedures in that scholars left Hebrew roots to Semitologists and turned to Indo-European. With the appearance of Brugmann and Delbrück's *Grundriß* (1897–1916), in principle, of Brugmann's part, a list of such roots acquired its canonical form. However, even then not everybody believed that the discovery of roots was the ultimate goal of etymology.

A great change was instituted by Per Persson (1912), who introduced the idea of extensions, or determinatives (*Erweiterungen*). His revolution can hardly be called an unmixed blessing, because for the first time ever historical linguists began to work with desemanticized units and confronted a high number of homonymous roots endowed with extremely vague meanings. The most cursory comparison of the lists in WP and *IEW* shows that Pokorny expunged whole pages from WP, and no one seems to have minded the loss. But Persson was a brilliant etymologist. Hundreds of his suggestions look plausible, and the determinatives allowed his followers to reach what looked like greater depths in the development of Indo-European.

A well-known reaction to Persson's approach was the counterrevolution in the form of a heightened interest in the "life of words". This laudable interest is easier to proclaim than maintain in its pure form. The parade example of *Histoire des mots* is EM. Despite its fame, EM is not an exemplary dictionary. Meillet, Persson's opponent from the start, wrote skimpy etymologies and too often made do with the formula "origin unknown", even when WH lists some conjectures worthy of note. German scholars also attempted to separate the two aspects and also with questionable success. Kluge (*EWDS*) concentrated on word origins, while Paul (*DW*) traced the words' history after they surfaced in the texts. Such was the initial plan. In principle, both remained true to their design, but, as could be expected, Kluge often discussed recorded history (especially when the word was fairly recent), while Paul could not always steer clear of etymology (especially when the word was old). The merger of etymology and history is even more noticeable in the subsequent editions of both

dictionaries. Most probably, the best etymological dictionaries of the future will combine information on the distant past (prehistory) and the period within human memory, but will succeed in striking a balance between the two.

The next step in pushing the temporal limits was connected with the reconstruction of the Indo-European laryngeals. The roots as we know them from IEW were projected to much greater antiquity, almost to the beginning of linguistic time. Fifteen years ago, Brill began to publish a series of etymological dictionaries, envisioned as preparatory studies for a new up-to-date dictionary of Indo-European that would eventually replace *IEW*. Despite the features uniting the members of the Leiden school (especially attention to the substrate and the laryngeals), each volume bears the imprint of its author's expertise and personality. Also, the tasks were different: some investigators had to deal with entire groups (Celtic and Germanic), others with separate languages, to say nothing of the different traditions underlying the assignments. Thus, Boutkan and Siebinga (2005) wrote the first etymological dictionary of Old Frisian. By contrast, Kroonen (2013) dealt with all the Germanic languages and could not even try to master most of the literature on such a subject (there is enough to read even on Old Frisian). The same holds for Latin and Classical Greek. Kroonen began his project when he was a graduate student and wrote the dictionary in two years, after getting his Ph.D. degree. The much more experienced Michel de Vaan completed the Latin dictionary (De Vaan 2008) in one year, a circumstance that aroused Blažek's admiration (Blažek 2007-2009/2013: 113) and leaves me overawed.

It would be unproductive and even unfair to compare such dictionaries with Feist's, Walde's (WP and WH), or even Levitsky's (2010; a more modest enterprise). In the preface, Kroonen admits his limitations. His work belongs to the same class as Orel's (2003). The Leiden dictionaries will be put to good use by the next editor of *IEW*'s successor, but they cannot be looked upon as models for the future. Their additional disadvantage consists in that their authors proposed many original solutions (acceptable or dubious, as the case may be), but the hayrick is huge, and finding those precious needles in them is next to impossible.

Just as, in my opinion, the profession needs a limited dictionary of words customarily dismissed as impenetrable, it needs a dictionary of reconstructed roots, laryngeals and all (with detailed discussion and references), and perhaps a dictionary of words of allegedly substrate origin. The last point deserves an additional comment, and one example will suffice. The indubitable and presumable cognates of Engl. *lie* 'tell falsehoods' are well-known (apart from the Germanic dictionaries, see *EESI* 16: 233–237 – Slavic). Despite such an abundance of attested forms, the verb's distant Indo-European etymology evades us. Discussion of the Gothic homonyms *liugan*¹ 'to lie' and *liugan*² 'to marry' and of the possible Baltic congeners of *liugan* has occupied researchers for almost two centuries. Boutkan and Siebinga (2005: 239, -*liuga*) state, perhaps rashly, that Germanic has related forms only in Old Church Slavonic and conclude: "…hence apparently a substratum word (pace Pokorny 1959: 686–7)." A dictionary of all the candidates for the substrate would make it possible to appraise such claims and perhaps allow us to arrive at less revolutionary results. As regards

roots, we already have Watkins (2000), but the material there is a rehash of *IEW*; it is a convenient but unoriginal compendium.

I would like to summarize my views on what kind of etymological dictionaries the future needs.

- 1. For some European languages good one-volume etymological dictionaries have already been written. There is no need to keep churning out their clones or slightly updated versions. The most typical example is English. As one can see from my overview (Liberman 1998), the post-Skeat titles are rather numerous. The production of those books, with the partial exception of Weekley (1921), was a waste. Even minor revisions of such dictionaries hardly contribute anything to scholarship and education. We can also look at German. When Seebold brought out the 22nd edition of "Kluge", one could not but welcome it, even though the gains accompanied the losses in the version he offered, but the next three editions (insignificant refurbishings) were not worth the effort. Lehmann's experience shows that even updating a dictionary of a dead language with a limited vocabulary (here Gothic) runs into almost insurmountable difficulties. No doubt, Lehmann read the post-1939 literature on his subject, but he had limited space at his disposal and as a result produced a useful supplement to Feist, rather than a new edition of the great work. A true revision should look like WH: in principle, a new dictionary.
- 2. From what has been said here it does not follow that such well-researched (even over-researched) languages as English do not need new etymological dictionaries. The opposite is true. "An English von Wartburg" is catastrophically overdue, minus of course a catalogue of forms through the centuries, because those can be found in the *OED*. Such a dictionary, clearly not limited to a thousand odd pages like *ODEE*, will offer a critical analysis of everything said about English words (rejecting, rather than dismissing unacceptable solutions; here Feist, with his peremptory *abzulehnen*, is not a good example). An undertaking of this type presupposes a considerable expense of time and money, but "the regime of stringent economy" in such matters will result in ignoring the achievements of the predecessors and supporting untenable conclusions.
- 3. An etymological dictionary of the future should not be used as a platform for airing the author's or authors' views on the structure of asterisked forms or the musings on the substrate and prehistoric migrations. A supplement on such matters will do.
- 4. It seems reasonable to let the public use the resources already in existence, target the scholarly community, and start publishing thematic etymological dictionaries. An explanatory volume of Indo-European or Nostratic roots belongs here. Assuming the existence of a database featuring the various opinions on the origin of words to be included, it will pay off to bring out a volume of words of presumably native origin about which there is something to be said. The next or the preceding volume can be devoted to the words of undiscovered origin.
- 5. Still another volume, sometimes the product of international cooperation, should deal with presumably borrowed words. For instance, in Finnish philology words

of Germanic origin loom large. In English, the main lender is Romance, especially French and Latin. Such words cannot always be discussed in depth, for Finnish is not Germanic and English is not Romance, but summaries and references should be taken for granted. My idea of concentrating on thematic volumes is not new, as evidenced by etymological dictionaries of verbs and adjectives. The latest (and, to my mind, highly successful) sample of this industry is Faltings (2010). If the plan I propose ever becomes reality, concise versions of multivolume dictionaries will follow. At present, society has almost abandoned funding basic sciences in the humanities, but only patience and understanding result in the production of *GDW*, *OED*, *FEW*, and similar masterpieces. Such projects exist even today. *DARE*, though completed and nearly bankrupt, still has something to do. *ESSI* and *EWA*, both admirable, are moving forward. If etymological dictionaries survive as a genre, those dictionaries should serve as models and shining examples of perseverance.

References

Blažek V. 2007–2009/2013. (rev.). De Vaan M. 2008. – *Philologica Finno-Ugrica* 13–15: 113–136. Boutkan D., Siebinga S. 2005. *Old Frisian etymological dictionary*. [Leiden Indo-European etymological dictionary series 1]. Leiden, Boston.

Brugmann K., Delbrück B. 1897–1916. *Grundriß der vergleichenden Grammatik der indo- germanischen Sprachen*. Berlin.

DARE = Cassidy F.G., Hall J.H. (eds.). 1985 – 2014. *Dictionary of American Regional English*. Cambridge (MS), London (UK).

De Vaan M.A.L. 2008. Etymological dictionary of Latin and the other Italic languages. [Leiden Indo-European etymological dictionary series 7]. Leiden, Boston.

DW = Paul H. 1987. *Deutsches Wörterbuch*. Halle a. S.

EM = Ernout A., Meillet A. 1932. Dictionnaire étymologique de la langue latine. Histoire des mots. Paris.

ESSI = Trubachev O.N. (ed.). 1974 - Etimologicheskii slovar' slavianskikh iazykov. Praslavianskii leksicheskii fond. Moscow.

EWA = Loyd A.L., Lühr R., Springer O. 1988-. Etymologisches Wörterbuch des Althochdeutschen. Göttingen, Zürich.

EWDS = Kluge F. 1884/1989/1995/2003/2011. *Etymologisches Wörterbuch der deutschen Sprache*. [22nd, 23rd, 24th, and 25th editions, by E. Seebold]. Straßburg, Berlin.

Faltings V.F. 2010. Etymologisches Wörterbuch der friesischen Adjektiva. Berlin, New York. Feist S. = VWGS.

FEW = von Wartburg W. 1934 – [1988]. *Französisches etymologisches Wörterbuch* (...). Leipzig, Berlin.

GDW = Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm. 1854 -1971. Leipzig. Hofmann J. = WH.

IEW = Pokorny J. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern, München.

Kennedy A.G. 1927. *A bibliography of writings on the English language from the beginning of printing to the end of 1922.* Cambridge (MS), New Haven, London, Oxford.

KEWA = Mayrhofer M. 1956 – 1980. *Kurzgefaßtes etymologisches Wörterbuch des Altindischen. A concise etymological Sanskrit dictionary.* Heidelberg.

Kluge F. = EWDS.

Kroonen G. 2013. Etymological dictionary of Proto-Germanic. [Leiden Indo-European etymological dictionary series 2]. Leiden, Boston.

Lehmann W.P. = VWGS.

Levitsky V.V. 2010. Etimologicheskii slovar' germanskikh iazykov. An etymological dictionary of Germanic languages. Etymologisches Wörterbuch der germanischen Sprachen. Vinnitsa.

Liberman A. 1992. Etymological studies IV. The 'dregs' of English etymology. – *General Linguistics* 32: 16–35.

Liberman A. 1998. An annotated survey of English etymological dictionaries and glossaries. – *Dictionaries* 19: 21–96.

Liberman A. 2003. The changing models of etymology. – Maggi D., Poli D. (eds.). *Modelli recenti in linguistica. Atti del Convegnio della Società Italiana di Glottologia, Macerata, 26–28 ottobre 2000.* Rome: 11–40.

Liberman A. [with the assistance of J. L. Mitchell]. 2008. *An analytic dictionary of English etymology: An introduction*. Minneapolis, London.

Liberman A. [with the assistance of A. Hoptman, N. E. Carlson]. 2010. A bibliography of English etymology (...). Minneapolis, London.

Mayrhofer M. = KEWA.

ODEE = Onions C.T. (ed.). [with the assistance of G. W. S. Friedrichsen, R. W. Burchfield]. 1966. *The Oxford dictionary of English etymology*. Oxford.

OED = Murray A.H. et al. (eds.). 1884–1928. The Oxford English dictionary. [3rd edition; www.oed.com]. Oxford.

Orel V. 2003. A handbook of Germanic etymology. Leiden, Boston.

Paul H. = DW.

Persson P. 1912. Beiträge zur indogermanischen Wortforschung. Uppsala, Leipzig.

Pokorny J. = IEW.

REW = Vasmer M. 1950 – 1958. Russisches etymologisches Wörterbuch. Heidelberg.

Seebold E. = EWDS.

Skeat W.W. 1910. An etymological dictionary of the English language. [4th edition]. Oxford.

Trubachev O.N. = ESSI.

Vasmer M. = REW.

VWGS = Feist S. 1939/1986. *Vergleichendes Wörterbuch der gotischen Sprache*. [3rd edition; 4th edition by W. P. Lehmann]. Leiden.

von Wartburg W. = FEW.

Watkins C. 2000. *The American Heritage dictionary of Indo-European roots*. [2nd edition, revised and edited by C. Watkins]. Boston, New York.

Weekley E. 1921. An etymological dictionary of Modern English. London.

WH = Walde A. 1938–1954. *Lateinisches etymologisches Wörterbuch*. [3rd edition by J. B. Hofmann]. Heidelberg.

WP = Walde A. 1927–1932. *Vergleichendes Wörterbuch der indogermanischen Sprachen*. [edited by J. Pokorny]. Berlin.