Etymology (A Linguistic Window onto the History of Ideas)

Version 1.0

June 2008

Joshua T. Katz

Princeton University

Abstract: This short essay for a volume on the classical tradition aims to give a basic, lively account of the forms and development of etymological practice from antiquity to the present day.

Written in the fall of 2007 and spring of 2008, this 3,000-word entry will be published (I hope more or less in its present form) in *The Classical Tradition*, a reference work from Harvard University Press edited by Anthony Grafton, Glenn Most, and Salvatore Settis.

***

In the first decades of the seventh century CE, Isidore, Bishop of Seville, compiled a 20-book work in Latin called *Etymologiae sive origines* ("Etymologies or Origins"). Our knowledge of ancient and early medieval thought owes an enormous amount to this encyclopedia, a reflective catalogue of received wisdom, which the authors of the only complete translation into English introduce as "arguably the most influential book, after the Bible, in the learned world of the Latin West for nearly a thousand years" (Barney *et al.*, 3). These days, of course, Isidore and his "Etymologies" are anything but household names—the translation dates only from 2006 and the heading of the wikipedia entry "Etymology" warns, "Not to be confused with *Entomology*, the scientific study of insects"—but the Vatican is reportedly considering naming Isidore Patron Saint of the Internet, which should make him and his greatest scholarly achievement known, if but dimly, to pretty much everyone.

People today are liable to confuse "etymology" with "entomology" because the words look and sound similar and, furthermore, because neither is so common, or describes so widespread a pursuit, as to be part of semantically transparent everyday discourse. Isidore himself would not have mixed them up: he knew Greek and understood that his subject was not "the study [-*logy*] of insects [*entomo-*]" but instead "the study of truth [*etymo-*]"—or, as he puts it, "the origin of words, when the force of a verb or a noun is inferred through interpretation" (1.29.1; trans. Barney *et al.*, 54). But it is not out of the question that he would nevertheless have believed them to be connected: perhaps there are *bees* in the *ABC's*? If it is of questionable judgment in a serious handbook to ascribe to Isidore, even in jest, an English-language-based case of wordplay (though I take comfort in knowing that the ardent wordsmith and lepidopterist Vladimir Nabokov shows himself scient in his 1969 novel *Ada* of the power potential in the anagrams *insect*, *incest*, and *nicest* and also that the most Isidorean of contemporary Latinists, John Henderson, has recently written a pun-filled book on the "Etymologies"), it is nonetheless appropriately illustrative, I believe, of the leading principle of ancient etymological practice, namely that things that sound even vaguely similar are the same in origin. For example, Isidore writes that "Fire [*ignis*] is so named because nothing can be born [*gignere*] from it, for it is an inviolable element, consuming everything that it seizes" (19.6.5; trans. Barney *et al.*, 376-377)—or, in Henderson's rendering, "'*Fire*' is so-called because *no* way to *sire* from it: . . ." (201). This is a case of an etymology "*e contrario*," that is to say, "from the opposite," a negative method that may well strike modern critics as peculiar. But catch as catch can, and it is noteworthy that Isidore's first-century BCE Roman forebear Varro took essentially the opposite tack in his etymology-filled work *De lingua Latina* ("On the Latin Language"): "Fire [*ignis*] is from being born [<g>*nascendo*] because from it there is birth, and fire sets alight everything that is born" (5.70).

Still, both Varro and Isidore employ the same basic etymological technique, deriving word *X* in language L from some other (like-sounding and either positively or negatively semantically connected, or at least connectable) word *Y* in the same L. Easily the best-known example of this method is *Lucus a non lucendo* "A grove [*lucus*] is so called from [*a*] not [*non*] being light [*lucendo*]": found as such in the commentary on Vergil (apropos of *Aeneid* 1.22) of

the fourth-century CE grammarian Servius (and in various other guises through the ages: see, e.g., Isidore 1.29.3, 1.37.24, 14.8.30, and 17.6.7), this catch phrase is now infamous for expressing an absurd idea, as though "A grove is so called from not being grave." And this in turn has given rise to the waggish *Ludus a non ludendo* "School [*ludus*] is so called because it's not cool [*ludendo*, literally 'playing']," a line whose first appearance in precisely this form seems only to be modern, though such major intellectual figures as Aelius Stilo (second/first century BCE; *frag.* 59 Funaioli), Varro (according to Isidore 18.16.2), Quintilian (first century AD; *Orator's Education* 1.6.34), and Festus (second century AD; p. 109, 23-24 Lindsay) report the essential idea (compare Maltby, 350). However, it was surely clear to many speakers of Latin that *ludus* actually *does* in some sense come from *ludendo*, for the basic meaning of the noun *ludus* in classical times was "play," not "(elementary or gladiatorial) school," a secondary sense that it probably acquired as a calque on Greek *skʰolḗ* "leisure → school." And as for *lucus*, some ancient sources report that it truly is *a lucendo* rather than *a non lucendo* (see Maltby, 349-350), a view with which, as it happens, modern etymologists concur: a clearing (the original meaning of *lucus*) is called thus from being clear.

To return to fire, the language L to which Varro's and Isidore's etymologies of *ignis* are specific is Latin, and it should not come as a surprise that in the very first work in the Western tradition devoted to the origin of language and to etymology, Plato's fourth-century BCE dialogue *Cratylus*, Socrates engages in the same procedure for Greek. When, for instance, Hermogenes asks Socrates to account for the Greek word for "fire," *pŷr*, the philosopher says that it does not easily fit with the Greek language, suggests that it is originally a foreign ("barbarian") expression, and claims that one would therefore be making quite a mistake if one tried to use Greek to explain its etymology (409C10-410A8)—but he leads off by saying, *Tò "pŷr" aporõ̂* "I have no idea about 'fire'" (409D1), a pithy phrase that deftly plays the usual linguistic game since the verb *a-porõ̂* ("I *fear* I don't know") looks to be the negation (in *a-*) of *pŷr* itself.

The Varronian etymology of *ignis*, cited above, explicitly involves derivation and thus has at least an implicit diachronic dimension: *X* comes from *Y* suggests that *Y* precedes *X*. But the ancient world also knew another technique, grounded in Epicurean and Lucretian atomism, that is today often referred to as etymological: likewise involving just a single language, it is above all playful (like, probably, Socrates' comment about *pŷr*), though it frequently has a serious, "scholarly," purpose as well (see, e.g., Snyder). In linguistic atomism, the smallest elements of language are the letters—known in Latin as *elementa* (a word whose etymology may well be the alphabetic sequence *LMN*)—and the complex arrangement of letters on the page is a major constituent feature of verse. When, therefore, the first-century BCE Roman poet Lucretius writes in his six-book hexametric poem *De rerum natura* ("On the Nature of Things"), *postremo in lignis cinerem fumumque uideri, / cum perfracta forent, ignisque latere minutos* ("finally, ash and smoke should be seen in wood when broken, and little fires should hide there"; 1.891-892), he is describing with artistic iconicity a scientific belief about the make-up of the world: the material "element" fire is evidently connected with wood (*ligna*), and this fact is mirrored in the Latin language, in which the five linguistic "elements" of the word for "fire," *IGNIS*, are literally "in pieces of wood," *in lIGNIS*. In fact (though this need not occupy us here), the precise nature of the analogy is disputed and the philosophical situation is evidently very complicated. Although Lucretius repeatedly associates the two words in Books 1 and 2 of his poem (on occasion he also associates *ignis* and the verb "be born," e.g., at 1.783-784), he mocks the fifth-century BCE pre-Socratic philosopher Anaxagoras for his belief that fire actually resides in

wood (we might say in English, "in *firs*"): Lucretius 1.891-892, just quoted, is a counterfactual assertion (ash and smoke "should" be seen but actually are not), and the poet elsewhere says, *et non est lignis tamen insitus ignis* ("and yet fire is *not* implanted in wood"; 1.901).

It is only later—much later—that people begin to etymologize across languages in a less scattershot way, in a way that would presumably be less likely to merit the scorn of such figures of the Enlightenment as Voltaire, to whom is attributed (perhaps incorrectly) the now infamous remark that etymology is the science in which consonants count for little and vowels for nothing. True, already in antiquity, learned men and poets would regularly compare or translate between Greek and Latin, or play on perceived connections (see, e.g., O'Hara), and some etymological ideas were probably commonplace: for example, that the words for "god" in these two languages, respectively *t^{h}eós* and *deus*, were somehow the same (see Maltby, 185). (This view seems so evidently correct that people are always surprised to learn that modern linguistics has conclusively demonstrated it to be wrong: neither do the two words go back to a common source nor does one derive from the other.) But "god" was a special case: early work from the Middle Ages and especially the Renaissance on linguistic kinship was generally intimately tied to one or another conception of the Tower of Babel and the relationships among languages and ethnic groups (see Borst). It is quite remarkable that Joseph Justus Scaliger in his *Diatriba de Europaeorum linguis* ("Diatribe on the Languages of the Europeans"), composed in 1599, went against the prevailing idea of immediately postlapsarian monolingualism by asserting that names of "god"—since they surely did not change over time—were a good way to determine linguistic relationship and that *t^{h}eós* and *deus* were far enough apart in sound that Greek and Latin, two languages that we can now prove are related, had to belong to different *linguae matrices*, or "wombs of language." (Scaliger states elsewhere, however, that *t^{h}eós* did give rise to *deus*.) The appearance of the following remark on "fire" in Erasmus' 1528 dialogue *De recta Latini Graecique sermonis pronuntiatione* ("On the Correct Pronunciation of Latin and Greek") is thus striking for its reach beyond both words for "god" and the canonical trio of sacred tongues, Greek, Latin, and Hebrew: "*Pŷr* is another word which . . . we [Dutch] pronounce wrong. The Germans, who have borrowed the word from Greek, pronounce it right except for changing the original smooth consonant for an aspirate [*Feuer*]. The Dutch change it to a *v* [*vuur*]" (947; trans. of M. Pope). Erasmus had no theory of linguistic relationship, and in fact he made a significant mistake in deriving Dutch *vuur* (X in language L) from Greek *pŷr* (Y in a *different* language M): the Germanic words for "fire" are cognate with *pŷr*—i.e., they all go back to the same source, a language now known as Proto-Indo-European—rather than borrowed from it. It was not until the end of the 18th century and the first half of the 19th that historical/comparative linguistics, in which whole sets of words are systematically compared, got off the ground. (The best, though idiosyncratic, book-length study of etymological practice over the past two hundred years is by Yakov Malkiel, who concentrates on the development of the Romance languages out of Latin.)

With the dawn of "scientific" etymology it became possible to compare X in language L and Y in language M and explain how both of them derive from *Z in language N, where N is a reconstruction of something we know existed but for which there are no actual linguistic records (a so-called proto-language) and where the asterisk indicates that Z is thus a reconstructed proto-form (see Watkins 2000). The name of "fire"—more exactly, names—is instructive in this regard. The evident similarities between Latin *ignis* and words for "fire" in the Indic, Baltic, and Slavic languages—e.g., Sanskrit *agní-*, Lithuanian *ugnìs*, and Old Church Slavic *ognĭ*—could finally be understood as pointing to the fact that six or so thousand years ago, they were all one

and the same thing, a Proto-Indo-European noun that we would now reconstruct as something like *$h_1og^{(w)}nis$. Similarly, Greek *pŷr*, though bearing no resemblance to *ignis*, is almost the same as *pir*, the word for "fire" in Umbrian, a language very closely related to Latin; together with forms in many other Indo-European languages—e.g., English *fire* itself, Armenian *hur* "fire," and Czech *pýř* "ashes," as well as words for "fire" in two branches of the family that were not known until the 20th century, Hittite (Anatolian) *paḫḫur* and Tocharian B *puwar* / A *por*— they lead us to reconstruct a proto-form *$peh_2w\underset{\circ}{r}$. (The Latin word *ignis* did not survive into the Romance languages, being replaced in Proto-Romance by the accusative case-form of Latin *focus* "hearth," whose meaning shifted metonymically to "fire." Despite appearances, then, such words as Italian *fuoco*, Spanish *fuego*, and French *feu* have nothing etymological to do with English *fire*, German *Feuer*, and Dutch *vuur*—or for that matter with Greek *pŷr*, which Joachim Périon in his 1555 treatise *De lingae Gallicae origine* ("On the Origin of the French Language") explicitly claimed as the ancestor of *feu*.) As for why the Proto-Indo-Europeans would have spoken of fire in two different ways, it is noteworthy that the words in the former set (*ignis*, etc.) are masculine while those in the latter (*pŷr*, etc.) are neuter. This distinction has suggested to researchers in the past century, both grammarians and those with a broader interest in stories and myths, that the one is an active force, capable of personification (Agni is the Vedic god of fire), while the other represents fire as an "inactive" natural substance. It may be worth noting in this regard that a linguist who has contributed a rather different entry for "Etymology" to another encyclopedia starts off his "History of etymology" by citing textual (rather than reconstructed) support that the earliest Indo-European people of the Indian subcontinent regarded Agni as a specifically active force: in the *Rigveda*, forms of the verb *aj-* "drive," the Sanskrit cognate of the like-meaning Latin *agere* (from whose past participle, *actus*, comes English *active*), are occasionally used in connection with the god (compare Hamp, 7). Even if there is a real basis for the link between Agni and *aj-* (and it must be admitted that the evidence is not as robust as one might wish), it is folk-etymological (see below) rather than scientific since there is no formal way to bring the noun *$h_1og^{(w)}nis$ (which may contain a root *h_1eg^w- "shine") together with the root *$h_2e\acute{g}$- "drive."

Now that a number of approaches to etymology have been mentioned, which yield results that are difficult to reconcile with one another (e.g., that Latin *ignis* is related to *gignere* and also to *ligna* and also to Sanskrit *agní-*), it is appropriate to come back to the literal definition of "etymology" as "the study of truth" and ask what this means. Are some of these approaches and results "more true" than others? And (to return to Isidore's definition) through what sorts of "interpretation" is it sensible to "infer" the "force" of a word? As so often with such questions, there are no simple answers: much depends on historical context and intellectual stance (compare Del Bello on postmodern etymological allegoresis). From classical times until very recently, as I have been trying to show, the vast majority of etymologies, whether rooted in one language or more, rely in the first place on the story one tells about similarities in sound; these days, however, linguists interested in etymology mostly compare across languages while looking backward at some proto-language, and the vagaries of phonological change mean that just as forms that look similar are often historically unrelated (e.g., Greek *tʰeós* and Latin *deus* or French *feu* and German *Feuer*), so do forms often turn out to be related that do not at first glance seem to have much to do with each other (e.g., the Latin root-cognate of Greek *tʰeós* is not *deus*, but rather *festus* "festive"). The linguist and classicist Roland G. Kent, editor of the standard Loeb Classical Library edition of *De lingua Latina*, comments in a terse footnote to Varro's claim that *ignis* is from <g>*nascendo* that this etymology is "[f]alse"—by which he means that

linguists today know that the Proto-Indo-European root of "be born" is $*\acute{g}enh_1$- and unconnected to $*h_1og^{(w)}nis$. This does not mean, however, that Varro's observation is worthless. Far from it: the Romans of Varro's time had no idea of Proto-Indo-European; more than a few of them will have arrived, however, at the idea that "fire" and "birth" had something to do with each other, and thoughts of this kind have subtle linguistic and cultural consequences. A well-known example of such a consequence is the common spelling of the name Publius Vergilius Maro not as "Vergil," but rather as "Virgil," a change due in part to the poet's reputation in the Middle Ages as a great magician or necromancer, someone who would have wielded a *uirga* (*virga*) "wand."

Broadly speaking, then, classical "etymology" in Late Antiquity (see Opelt), the Middle Ages (see Klinck), and still beyond was largely a matter of what is today retronymically referred to as "folk (or, popular) etymology." But folk etymology, though often derided by those "in the know," remains an important linguistic force and must be taken seriously, for one thing because the popular form often wins out. For example, the *mouse* in our word *dormouse* (plural *dormice*) reflects the influence of murine creatures on (probably) an Anglo-Norman word like *dormeus* "hibernating" (compare the sleepy Dormouse in *Alice's Adventures in Wonderland*); *rhyme* is hardly ever written *rime* any longer (as in "The Rime of the Ancient Mariner"), but the spelling with *rhy-*, which strongly suggests a borrowing of a rho-initial word in Greek (a language that, in the classical period, did not use rhyme as a poetic device), shows the influence of the (in fact unrelated, Greek-derived) word *rhythm*; and the Greek-derived word *asparagus* has since the 17th century often been called *sparrow-grass* (indeed, the *Oxford English Dictionary* s.v. "asparagus" notes that *sparrow-grass* "remained the polite name during the 18th c[entury, though b]otanists still wrote *asparagus*").

Isidore writes that "The knowledge of a word's etymology often has an indispensable usefulness for interpreting the word, for when you have seen whence a word has originated, you understand its force more quickly. Indeed, one's insight into anything is clearer when its etymology is known" (1.29.2; trans. Barney *et al.*, 55). Since language is an integral part of the human experience, the species-unique tool we use to describe and order our surroundings, it is understandable that people should wish to examine the tool itself: to take it apart, to play with it, to try to square the word with the world. Long may we continue to do so—and there are ample classical and post-classical models at our disposal.

Bibl.: Stephen A. Barney, W. J. Lewis, J. A. Beach, and Oliver Berghof, *The* Etymologies *of Isidore of Seville* (Cambridge, 2006). Arno Borst, *Der Turmbau von Babel: Geschichte der Meinungen über Ursprung und Vielfalt der Sprachen und Völker*, 4 vols. (Stuttgart, 1957-1963). Davide Del Bello, *Forgotten Paths: Etymology and the Allegorical Mindset* (Washington, D.C., 2007). Eric P. Hamp, "Etymology," in *International Encyclopedia of Linguistics*, 2nd ed., ed. William J. Frawley, vol. 2 (Oxford, 2003): 7-12. John Henderson, *The Medieval World of Isidore of Seville: Truth from Words* (Cambridge, 2007). Roswitha Klinck, *Die lateinische Etymologie des Mittelalters* (Munich, 1970). Yakov Malkiel, *Etymology* (Cambridge, 1993). Robert Maltby, *A Lexicon of Ancient Latin Etymologies* (Leeds, 1991). James J. O'Hara, *True Names: Vergil and the Alexandrian Tradition of Etymological Wordplay* (Ann Arbor, 1996). I. Opelt, "Etymologie," in *Reallexikon für Antike und Christentum*, vol. 6, ed. Theodor Klauser (Stuttgart, 1966): cols. 797-844. Jane McIntosh Snyder, *Puns and Poetry in Lucretius'* De rerum

natura (Amsterdam, 1980). Calvert Watkins, *The American Heritage Dictionary of Indo-European Roots*, 2nd ed. (Boston, 2000).