Etymology

Steven N Dworkin, University of Michigan, Ann Arbor, MI, USA

Abstract

This article begins by defining etymology as both the identification of word origins and the preparation of histories of individual words or word families. It next offers a brief overview of the history of this linguistic subdiscipline. The article then discusses such topics as the relationship of etymology to other branches of historical linguistics, the diffusion of the findings of etymological research (especially in etymological dictionaries), theoretical and methodological questions, and the insights provided by etymological reconstruction of the lexicon of undocumented language states into the cultures and social structures of past historical periods.

Definition and Scope of Etymology

Etymology can be concisely defined as the search for and identification of the origin of a given word in a specific language or of a given series of historically related words in a language family. It is the etymologist's responsibility to uncover and describe the morphophonological form and the semantic range of the base that forms the starting point for the evolution over time and space of the lexical item under investigation. For many specialists in this field, etymology has come to designate the preparation of the complete history or biography of a word. Identification of the word's immediate formal and semantic origin is now but the first step in this task, which includes accounting for all facets of the chosen word's formal and semantic history, the derivatives to which it has given rise, its rivalry with other referentially similar lexical items, its spread into other languages as a loanword (if appropriate) and, when relevant, the circumstances leading to its demise or slide into obsolescence. This approach to etymology seems feasible only for languages with a long and abundant recorded history. Etymology is consequently an integral part of the broader discipline of diachronic lexicology.

In general terms, a given word can (1) form part of the language's inherited lexicon, (2) represent a borrowing from a language with which the language under study has come into contact, or (3) result from the various (often language-specific) internal processes of lexical creation available to speakers, such as prefixal and suffixal derivation, infixation, compounding, sound symbolism, and lexical blends. The identification of the elements that fall into each of these three broad categories in a specific language forms part of the brief of etymology, the sole linguistic discipline that is exclusively diachronic in scope. Most etymological research devoted to individual languages or language families concentrates on the history of the inherited lexicon, and, to a lesser extent, on loanwords. Derivatives have traditionally received far less attention. Yet, the history of each apparent derivative deserves individual study in order to differentiate authentic derivatives created within the language at issue from those which, genetically, are actually inherited from the lexicon of the ancestor language, or are borrowings from yet another language (e.g., the many Spanish derivatives in – *aje* which, for the most part, are

borrowings from neighboring Gallo-Romance and Catalan). Etymology is one branch of historical linguistics that needs to make use of the findings of many of the other fields, such as phonology, inflectional and derivational morphology, syntax, semantics, and lexical typology, as well as relevant extralinguistic knowledge garnered from the study of political, social, cultural, and literary history, and textual criticism. Unless the analyst is dealing with the lexicon of an isolate language (such as Basque), research into the history of the lexicon of an individual language needs to be comparative in nature, with the presentation of cognates from other members of the relevant language family. Often, evidence from a related language may tip the scales in favor of one of several competing etymological hypotheses, or even point the way to a new solution. Even when dealing with an isolate language, comparative evidence from attested regional varieties of that language may be useful in etymological reconstruction.

Etymology in the Past

The origin of words has attracted the attention of scholars and thinkers interested in human language for over two millennia. Early philosophers thought that the original meaning of a word reflected its true essence, a belief reflected in the Greek label *etymologia* (a compound of *etymos* 'true' and *logos* 'word'). Such etymologizing plays a significant role in various religious traditions. The Bible and the subsequent Jewish and Christian traditions of biblical exegesis contain numerous passages in which the text seeks to identify the original essential meaning of a word, especially of personal and place names. A similar approach to the relationship between linguistic signs and the realities that they name is found in the Chinese *Shih ming* (*c*.CE 200). Old French *ethimologie* and Old Italian *etimologia* are documented in thirteenth-century texts. English *etymology*, rarely found before the sixteenth century, is a borrowing from Old French. With the awakening of scholarly interest in vernacular languages during the Renaissance, there began to appear in the sixteenth and seventeenth centuries, etymological dictionaries of numerous European languages. Many of the etymologies proposed in these prescientific works are fanciful, often based on fortuitous formal and semantic resemblances.

In the western scholarly tradition, etymology became a truly scholarly discipline only with the establishment of the comparative method and the founding of Indo-European (IE) comparative linguistics in the early nineteenth century with its emphasis on the regularity of sound change, a basic underlying principle of etymological research and reconstruction. This period witnessed the publication of the first great etymological dictionaries of IE languages (e.g., Pott, 1833–36; Diez, 1853; Fick, 1871; Kluge, 1883). As a scholarly discipline, etymology forms part of the European linguistic tradition. The major etymological dictionaries of numerous non-European languages and language families were prepared over the course of the last century by European and American experts.

Etymology and Other Linguistic Fields

Etymology finds itself in a symbiotic relationship with some of the other disciplines of historical linguistics. In 1875, the American pioneer William Dwight Whitney (1827–94) declared (echoing thoughts voiced earlier in the century by Jakob Grimm): "The whole process of linguistic research begins in and depends upon etymology" (quoted in Malkiel, 1993: 20). The sound correspondences that form the backbone of the study of a language's formal evolution are extrapolated from a large quantity of accepted and verified etymologies. An equally reliable etymological database is needed to study the patterns of semantic evolution. Yet, a deep knowledge of formal and semantic evolutionary patterns identified on the basis of well-established etymologies is essential for the reconstruction or identification of the etymon of the many words in a language whose origins are unclear, controversial, or completely unknown. Many of the nineteenth-century pioneers of etymological research authored both historical grammars and, separately, etymological dictionaries of the same language or language family, for example, Jakob Grimm (1822–1837), 1854–1960 (the latter work was carried out in collaboration with his brother Wilhelm Grimm)) for German and Friedrich Diez (1836–1844, 1853) and Wilhelm Meyer-Lübke (1890–1902, 1911–1920) for Romance. In all three cases, publication of the historical grammar preceded the corresponding etymological dictionary. The close relationship between etymology and other fields of historical linguistic research constitutes one of the main themes running through the etymological writings of the Romanist Yakov Malkiel (see, for example, the essays collected in Malkiel, 1968).

Etymology has to distinguish between the immediate and the distant origin of the word under study. Etymological studies of the lexicon of a specific language tend to work backward from the modern language. Historians of each language or language family must delimit, in light of the available linguistic and extralinguistic data, how far back in time to trace the word's origin. For example, etymological studies of the inherited lexicon of a Romance language traditionally seek to identify the appropriate (documented) Latin base of a Romance word, without discussing the more distant Italic or IE origin of the relevant Latin form (although there are exceptions, such as the Indo-Europeanist Giacomo Devoto's *Avviamento alla etimologia italiana*, which identifies, when relevant, the IE root of the etymon). In the case of the numerous borrowings from other languages that entered each Romance language through contact situations, the etymologist will identify the etymon that entered from the immediate source language, without elaborating on its origins in the donor language. To give but one example, the historian of the Spanish lexicon will identify Spanish *azul* 'blue' as a borrowing from Arabic (its immediate origin) but will not discuss the circumstances of this word's entry into Arabic from Persian. On the other hand, the author of an etymological dictionary of Gothic can choose to identify either the proto-Germanic or the proto-Indo-European root of the lexical item under examination.

Etymology and Reconstruction

On the basis of cognate sets, the oldest available data, and the application of the comparative method, etymological dictionaries for most languages or language families strive to reconstruct undocumented roots as the appropriate starting points. Practitioners of Romance etymology are blessed (or cursed?) by having at their disposal the rich documentation afforded by written Latin, reflecting the different written registers of the spoken language (labeled variously as Spoken Latin, Vulgar Latin, and Proto-Romance) underlying the Romance languages. Traditional Romance etymology has contented itself with presenting as the etymon the form of the base as it appears in written Latin, with asterisked hypothetical bases invoked as a last resort in the absence of a written Latin correlate. Often, the standardized orthography of the traditional Latin form masks the phonetic reality of the etymon in the spoken language, which can be recovered through application to the Romance data of the comparative method. This is the underlying principle of the *Dictionnaire Étymologique Roman* project codirected by Eva Buchi and Wolfgang Schweickard, in which the Proto-Romance starting point is presented in phonemic transcription (although the Romance descendants are offered in standard orthography). Each entry ends with explicit reference to the relevant form of the base in written Latin. I shall offer one example of this procedure. The Romance data permit the reconstruction at the level of Proto-Romance of the bases */askUlt-a-*/eskult-a-/, which correspond to ausculty */eskult-a-/are 'to listen' of written Latin. This use of comparative reconstruction in Romance etymology has been the subject of a lively debate among specialists (see, for example, Varvaro, 2011a,b and, in response Buchi and Schweickard, 2011a,b).

The attempt to identify the etymologies for a language whose genetic affiliation is uncertain poses different problems. Well-established cognate sets play an important role in the reconstruction of etymological bases. In the case of very large language families such as Sino-Tibetan Burman or Austronesian, whose genetic subgroupings are highly controversial, the etymologist must take care in weighing the role of possible cognates in an etymological reconstruction, and must distinguish between formal and semantic resemblances that reflect actual genetic affiliation versus those that may result from processes of language contact.

Over the years etymologists have refined their techniques for reconstructing the form of the relevant bases underlying the words or word families under study. The semantic side of

the (reconstructed) base has not been a primary focus of etymological research. Emile Benveniste's seminal study (1954) lays out many of the difficult and complex issues involved in semantic reconstruction. Typical is the declaration made in Orel (2003: xii): "The semantic reconstruction of the proto-Germanic words was not even attempted: too many complications and arguments would have followed the decision to ascribe meanings to Proto-Germanic words." Only recently have some specialists attempted to apply the principles of the comparative method to the reconstruction of the semantic range of the proposed etymological base. It would be reasonable to assume that meanings shared by all or many of the descendants of a given base (reconstructed or documented) go back to the semantic range of that item in the protolanguage. Meanings found only in a small number of geographically scattered cognate forms may reflect inheritances from the protolanguage, or, alternatively, might result from independent but parallel semantic innovations brought about by the working of such cognitive processes as metaphor and metonymy. It is certainly realistic to posit polysemy at the level of the etymon (Rankin, 2003). One of the innovative features of the aforecited *Dictionnaire Étymologique Roman* is its meticulous reconstruction and systematic ordering in each entry of the semantics of the posited proto-Romance base. The reconstructions are based on the attested semantic range of each Romance cognate rather than on the meanings of the written Latin equivalents.

Etymological Dictionaries

Etymological dictionaries of entire language families can adapt a prospective perspective, taking as the starting point for each entry the documented or, more often, reconstructed lexicon of the source language, and tracing the evolution of each base into the individual descendant languages. This approach is also feasible for the study of the vocabulary of individual languages. In addition to the pan-Romance *Romanisches Etymologisches Wörterbuch* (1911, 1920₃), the multivolumed *Französisches Etymologishes Wörterbuch* (von Wartburg et al., 1922–2002) and the ongoing *Lessico Etimologico Italiano* (Pfister, 1979) are arranged alphabetically by etymon. A similar structure is found in the innovative and methodologically controversial *Dictionnaire Étymologique Roman*, whose entries are arranged according to the reconstructed proto-Romance base (presented in phonemic transcription), rather than by the written Latin equivalent. This dictionary has limited its scope in its first stage of elaboration to approximately 500 proto-Romance bases that have survived into all, or most of, the Romance languages. However, this prospectively arranged system poses serious problems when deciding in which entry to place the many lexical items in the target language whose origins are far from transparent. The team responsible for the continuation of the *Französisches Etymologishes Wörterbuch* has had to prepare several fascicules, arranged by semantic categories, devoted to words of unknown origin. In dictionaries arranged according to the etymon, indexes of the words studied are essential for the nonspecialist reader interested in the origin of a specific word.

Etymological information can be transmitted in several forms. The layperson seeking to know the origin of a given word in a specific language will often find a concise identification of the etymology in the relevant entry in a monolingual reference dictionary. For example, most dictionaries of, say, French will tell the reader of the entry for *père* 'father' that the word comes from the Latin *pater*, but will not explain the formal or semantic details of its evolution. With regard to the origin and history of words whose etymology is unknown or disputed, specialists usually pass on their findings or hypotheses to their colleagues in the form of brief notes or more elaborated articles in scholarly journals, usually read only by specialists. Far less frequent today are monographs or even books devoted to the detailed study of the history of individual words or of entire word families. The culmination of etymological research on the lexicon of a language or language family is the etymological dictionary where each entry discusses, insofar as is feasible, the given word's formal and semantic history, as well as offering, when appropriate, a scholarly justification for the proposed etymology and, if appropriate, a critical review of competing explanatory hypotheses. Indeed, it is often an etymological dictionary whose entries make available to the wider scholarly community the findings first reported in detailed and specialized etymological articles and monographs.

Today, most major languages or language families have at least one etymological dictionary devoted to their lexicon. An excellent and thorough critical overview (with bibliography of primary and secondary sources) of such dictionaries appears in Eva Buchi's contribution to the forthcoming *Oxford Handbook of Lexicography*. Malkiel's (1976) typology of etymological dictionaries retains its value. He adduced eight criteria for classifying such dictionaries: (1) time depth (period to which the etymologies are traced back), (2) direction of analysis (prospection or retrospection), (3) range (languages dealt with), (4) grand strategy (structural division of the dictionary), (5) entry structuring (linear presentation of the chosen features), (6) breadth (information given in the front- and back matter vs within the individual entries), (7) scope (general lexicon vs parts of it, e.g., borrowings), and (8) character (author's purpose and level of tone). The entries in most such dictionaries are arranged alphabetically by the word whose origin it seeks to elucidate or by the chosen etymon, rather than by semantic categories or lexical families.

Etymology, Theory, and Methodology

For the most part, etymologists prefer to devote their efforts to the identification of etyma and the preparation of word histories/lexical biographies rather than to discussions of methodology and theory. Although Jules Gilliéron originally coined his famous dictum 'Each word has its own history' as a response to the Neogrammarian view of the rigidity of the so-called sound laws, his words could equally apply to lexical history, including etymology. Each etymological solution is unique. Consequently, etymology is a linguistic subdiscipline that does not readily lend itself to rigid methodological systematization or vociferous and polemical theoretical debate. There can be no unified theory of etymological analysis, a situation that may account for the current estrangement between etymology and general historical linguistics. Some renowned etymologists (e.g., Vittorio

Bertoldi have gone so far as to describe etymology as an 'art' (cf the title of his introductory manual L'arte della etimologia (1952)).

Etymology no longer occupies a central place in the discourse of historical linguistics. Etymological method today seldom receives lengthy attention in manuals and handbooks of historical linguistics and is rarely afforded separate book-length treatment. Notable recent exceptions in the English-speaking world are the overviews of etymology by Anatoly Lieberman (2005) and Philip Durkin (2009) aimed at the educated nonspecialist reader. At the beginning of the twentieth century, Romance etymologists debated the relative weight to be accorded to formal versus semantic criteria in resolving complicated etymological cruxes. In retrospect, this was a pointless discussion, as the etymologist must weigh all the relevant linguistic (and, if appropriate, nonlinguistic) evidence. There exists no algorithm for evaluating each case. Etymology is language specific, and most etymologists specialize in the history of one language or one language family. It is next to impossible to identify etymological universals. Can we even speak of 'general etymology' in the same way as 'general phonetics' or 'general semantics?' A comparative etymological analysis of the lexicon of individual languages or language families undertaken by a team of specialists may identify certain semantic fields or grammatical categories that possibly contain a high number of elements of unclear origin, or may reveal certain cross-linguistic patterns in lexical history. In the Romance languages, the origin of many of the words corresponding to 'boy, child' is controversial: Fr. *garçon*, Sp. *muchacho*, *niño*, *chico*, *mozo*, Ptg. *rapaz*, It. *ragazzo*, and Rum. *băiat*. Although outside the Romance domain, one may include here the English *boy*, originally 'servant.'

Etymology, Loanwords, and History

All words evolve in a historical, cultural, and social context. For many etymologists (including the author of this article), the opportunity to study lexical history in such a framework is one of the joys of humanistically slanted diachronic lexical research. The link between diachronic lexical studies and material reality can be seen in the *Wörter und Sachen* school of dialect geography that flourished among European linguists in the first half of the twentieth century. Such studies focused, from a historical perspective, on how a language or group of languages expressed a given reality of material civilization (e.g., animal-, plant-, bird-names, farm implements, tools, etc.). The identification and analysis of loanwords resulting from language contact (at the level of both oral and written language) require close examination of the relevant extralinguistic circumstances, including accurate dating and philological analysis of the first documentation of the relevant forms in both the source and recipient language. A loanword can exist in the recipient language for a long time before it surfaces in a text, the only source of documentation for past stages of the language. The historical study of loanwords requires the etymologist to study the integration and diffusion of the neologism through the speech community, processes that may take a long time after its introduction into the host language (for an example of the application of this approach to Spanish, see Dworkin, 2012). A recent cross-linguistic study of loanwords in about 40 languages (Haspelmath and Uri Tadmor, 2009) has demonstrated that lexical items belonging to certain semantic fields (e.g., emotions and values, motion, kinship, spatial relations, body parts, and sense perception), semantic areas that are universal to all human societies, are rarely borrowed. In contrast, fields such as clothing and grooming, religion and belief, the house, law, agriculture and vegetation, social and political relations, food and drink, domains subject to a high degree of intercultural influence, show higher rates of borrowing. These considerations may guide etymologists in deciding in doubtful cases whether it is reasonable to propose that a specific word may be a borrowing. The results of etymological research form the foundation of a subfield known as linguistic paleontology, whose findings allow historians to analyze (more often than not, reconstructed) lexical data in order to recover details or formulate hypotheses concerning early (often prehistoric) societies, their religious beliefs, their kinship systems, the structure of the local economy, etc. (cf the attempts to reconstruct diverse facets of proto-Indo-European society, or the attempts to use specific items of the reconstructed lexicon to locate the homeland of speakers of proto-Indo-European).

See also: Comparative Method in Linguistics; Indo-European Languages; Internal Reconstruction in Linguistics; Lexical Semantics; Lexicology and Lexicography.

Bibliography

Benveniste, E., 1954. Problèmes sémantiques de la reconstruction. In: Benveniste, E. (Ed.), Problèmes de linguistique générale, Word, vol. 10, pp. 131–144. Martinet, A., Weinreich, U. (Eds.), Linguistics today. Linguistic Circle of New York, New York. Reprinted in Benveniste, E., 1966. Problèmes de linguistique générale. Gallimard, Paris, pp. 289–307.

Bertoldi, V., 1952. L'arte dell'etimologia. Liguori, Naples.

Buchi, E. Etymological dictionaries. In: Durkin, P. (Ed.), The Oxford Handbook of Lexicography. Oxford University Press, Oxford, forthcoming.

Buchi, E., Schweickard, W., 2011a. Sept malentendus dans la perception du DERom par Alberto Vàrvaro. Revue de linguistique romane 75, 305–312.

Buchi, E., Schweickard, W., 2011b. Ce qui oppose deux conceptions de l'étymologie romane. Réponse à Alberto Vàrvaro et contribution à un débat méthodologique en cours. Revue de linguistique romane 75, 628–635.

Devoto, G., 1966. Avviamento alla etimologia italiana. F. le Monnier, Florence.

Diez, F., 1836–1844. Grammatik der romanischen Sprachen, 3 vols. Weber, Bonn.

Diez, F., 1853. Etymologisches Wörterbuch der romanischen Sprachen, fifth ed. Adolph Marcus, Bonn. 1887.

Durkin, P., 2009. The Oxford Guide to Etymology. Oxford University Press, Oxford.

Dworkin, S.N., 2012. The History of the Spanish Lexicon: A Linguistic Perspective. Oxford University Press, Oxford.

Fick, R., 1871. Vergleichendes Wörterbuch der indogermanischen Sprachen. Vandehoeck and Ruprecht, Göttinge.

Grimm, J., 1822–1837. Deutsche Grammatik. Bei Deitrich, Göttingen.

Grimm, J., Grimm, W., 1854. Deutsches Wörterbuch. S. Hitzel, Leipzig.

Haspelmath, M., Uri, T., 2009. Loanwords in the World's Languages: A Comparative Handbook. Mouton de Gruyter, Berlin.

Kluge, F., 1883. Etymologisches Wörterbuch der deutschen Sprache. K.J. Trübner, Strassburg. Numerous revised editions and reprints.

Lieberman, A., 2005. Word Origins and How We Know Them. Etymology for Everyone. Oxford University Press, Oxford.

Malkiel, Y., 1976. Etymological Dictionaries: A Tentative Typology. University of Chicago Press, Chicago and London.

Malkiel, Y., 1968. Essays on Linguistic Themes. Blackwell, Oxford.

Malkiel, Y., 1993. Etymology. Cambridge University Press, Cambridge.

Meyer-Lübke, W., 1896–1902. Grammatik der romanischen Sprachen, 4 vols. Fues, Leipzig. French Translation (1890–1906). Grammaire des langues romanes. 4 vols. Welter, Paris.

Meyer-Lübke, W., 1911–1920. Romanisches Etymologisches Wörterbuch, third ed. Winter, Heidelberg. 1935.

Orel, V., 2003. A Handbook of Germanic Etymology. Brill, Leiden and Boston.

Pfister, M., 1979. Lessico etimologico italiano. Dr. Ludwig Reichert, Wiesbaden.

Pott, A., 1833–36[1859–1876]. Etymologische Forschungen auf dem Gebiete der Indogermanischen Sprachen. Meyer, Lemgo.

Rankin, R., 2003. The comparative method. In: Joseph, B., Janda, R. (Eds.), The Handbook of Historical Linguistics. Blackwell, Oxford, pp. 183–212.

Varvaro, A., 2011a. Il DERom: un nuovo REW? Revue de linguistique romane 75, 297–304.

Varvaro, A., 2011b. La 'rupture épistemologique' del DERom: Ancora sul metodo della etimologia romanza. Revue de linguistique romane 75, 623–627.

von Wartburg, W., et al. (Eds.), 1922–2002. Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes, 25 vols. Klopp/Winter/Teubner/Zbinden, Bonn; Heidelberg; Leipzig-Berlin; and Basel.