

**Durkin, Philip (ed.). *The Oxford Handbook of Lexicography*.  
Oxford: Oxford University Press.**

### **Part III: Specialist dictionaries**

#### **Chapter 20: Etymological dictionaries**

##### **20.1. Introduction**

No other linguistic subfield is as closely linked to lexicography as etymology<sup>1</sup>. Indeed, whilst significant work on synchronic lexicology is done without any reference to dictionaries, major etymological breakthroughs, be they factual or methodological, are mostly expressed through lexicographic work, and when they are not, it is their subsequent acceptance by a reference dictionary which ultimately lends them support. Similarly, I know of almost no outstanding etymologist of our time who would not in some way be linked to a major lexicographic enterprise: most of them are either authors of completed or ongoing etymological dictionaries or current or former heads of etymological teams for general dictionaries.

However, if the strong relevance of etymological lexicography (or etymography) for scientific knowledge building is self-evident, there exists probably no general agreement about its scope. I follow here the definition Hartmann's and James' *Dictionary of lexicography (DLex)* gives of etymological dictionaries: "a type of DICTIONAR[IES] in which words are traced back to their earliest appropriate forms and meanings", this tracing back being their assumed principal

purpose. This means that general and/or historical dictionaries (for which see part II: Historical dictionaries, in particular chapter 14: The role of etymology and historical principles, as well as Schweickard 2011) will not be tackled here, although some of them, like the *Oxford English Dictionary (OED)* or the *Trésor de la langue française (TLF)*, contain encapsulated in them the best available etymological dictionary of the language they describe.

The element *word* in the *DLex* definition, although intuitively comprehensible, lacks technical rigour, and is therefore ambiguous. I will thus ban *word* from this chapter and make use instead of the threefold terminology (as well as the typographical conventions attached to it) established within the theoretical framework of Meaning-text theory (see Mel'čuk 2012: 1: 21-44): *wordform* (defined as 'segmental linguistic sign that is autonomous and minimal, i.e., that is not made up of other wordforms'), *lexeme* ('set of wordforms, and phrases, that are all inflectional variants'), and *vocable* ('set of lexical units –lexemes or idioms– whose signifiers are identical, whose signifieds display a significant intersection, and whose syntactics are sufficiently similar'). I find this terminology particularly useful for etymological and etymographical purposes: first because it is coherently based on Saussure's definition of linguistic signs and secondly because it reserves a term (*lexeme*) for the central unit 'one signifier, one signified,

all inflectional variants' of a polysemous vocable, which in most terminologies is not explicitly named (mostly, there is talk about "words" developing new "senses", but *sense* only refers to the signified and not to the combination of the signifier, the signified, and the syntactics)<sup>2</sup>. Thus, for example, the vocable TABLE –if one agrees, for the sake of simplicity, on describing TABLE as a (very) polysemous unit rather than as a set of homonymous ones– contains lexemes like TABLE1 'article of furniture consisting of a flat top and legs', TABLE2 'arrangement of items in a compact form', and TABLE3 'upper flat surface of a cut precious stone', which in turn present the wordforms *table* and *tables*; in general, dictionary entries are made up of vocables like TABLE.

A firm believer in the concept of proper names as a scalarly stratified part of the lexicon (see van Langendonck 2007), I nevertheless exclude here discussion of etymological dictionaries of place names (for which see chapter 15), personal names (chapter 16), and other proper names.

## **20.2. Contemporary practices in etymographical work**

Malkiel (1976) offered us a book-length typology of etymological dictionaries, analyzing them on the basis of eight autonomous criteria: (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). Amongst these criteria, I will use scope in order to distinguish not so much among different types of etymological dictionaries (although that will also be the case), but among three grand etymological classes, which each make their own different demands of an etymologist, and which are sometimes dealt with in different dictionaries: inherited lexicon (20.2.1.), borrowings (20.2.2.), and internal creations (20.2.3.). For each of these classes, I shall try to give a general idea of the (methodological) state of the art, mostly on the basis of etymological dictionaries of European languages, and to draw attention to what I take to be the most profitable approaches within the field.

### **20.2.1. Inherited lexicon**

Amongst the three major etymological classes, inherited lexicon clearly gets the most attention in terms of etymological dictionaries devoted to its study. One defining feature of this kind of etymological dictionary is its comparative character (see Forssman 1990 and Malkiel 1990: 1329-1330). Indeed, as the inherited lexicon is typically etymologized by comparative reconstruction, whole language families (or branches of them, also called families) are usually taken into consideration. As a

consequence, the arrangement of these dictionaries is prospective rather than retrospective (Malkiel 1976: 25-27), i.e. their lemmata pertain to the reconstructed protolanguage rather than to the individual languages on which the comparison is based. Usually, the underlying question these dictionaries set out to answer is where the inherited lexicon of currently spoken languages comes from, and their ultimate goal is to reconstruct the lexicon of a proto-language.

This is typically the case of the *Dictionnaire Étymologique Roman (DÉRom)*, which aims to reconstruct Proto-Romance, i.e. the common ancestor of the (spoken) Romance languages, following Jean-Pierre Chambon's claim that Romance etymology could benefit from the comparative method (see Chambon 2010). In this dictionary, comparative reconstruction is used, for instance, in order to reconstruct Proto-Romance \*/batt-e-/ trans.v. 'to beat' from Italian BATTERE, French BATTRE, Old Spanish BATER and their cognates (Blanco Escoda 2011/2012 in *DÉRom* s.v. \*/batt-e-/). What is standard practice in other linguistic domains is however quite unusual in the field of Romance etymology, where scholars usually discard the comparative method as unnecessary in the face of all the written testimonies of (mostly classical) Latin. The entries corresponding to \*/batt-e-/ in the three major reference dictionaries of Romance etymology, Meyer-Lübke's *Romanisches Etymologisches Wörterbuch (REW)*, von

Wartburg's *Französisches Etymologisches Wörterbuch (FEW)*, and Pfister's *Lessico Etimologico Italiano (LEI)*, are indeed made up of written items as found in Latin dictionaries: *battuëre* (REW 1935<sup>3</sup> [1911<sup>1</sup> : *battuere*]), *battuere* (von Wartburg 1924 in *FEW* 1, 290b), and *batt(u)ere* (Calò/Pfister 1995 in *LEI* 5, 344a). Currently, there is no agreement about the relevance of comparative grammar for Romance etymology (pro: Buchi 2010a and Buchi and Schweickard 2011; contra: Kramer 2011 and Vårvaro 2011): the methodological principles on which the *DÉRom* is based constitute an ongoing debate.

With the Indo-European Etymological Dictionary project of Leiden University (see *Indo-European Etymological Dictionaries Online, IEDO*), reconstruction goes even a step further and becomes articulated in a most interesting way: first, each of the etymological dictionaries of individual branches of Indo-European<sup>3</sup> reconstructs the inherited lexicon of their immediate protolanguage, which then enables reconstruction of the Proto-Indo-European lexicon. For instance, the *Etymological Dictionary of Latin and the other Italic Languages (IEEDLatin)* reconstructs, based on Latin, Faliscan, Oscan, Umbrian, and South Picene cognates, Proto-Italic *\*mātēr*, *\*mātr-* f.n. 'mother'. For its part, the *Etymological Dictionary of the Slavic Inherited Lexicon (IEEDSlavic)* uses Church Slavic, Russian, Czech, Polish, Serbo-Croatian, Čakavian, and Slovene cognates for

reconstructing Proto-Slavic *\*màti* f.n. ‘mother’. In the same way, the *Etymological Dictionary of Proto-Celtic (IEEDCeltic)* reconstructs from cognates from Irish, Welsh, Breton, Cornish, Gaulish, and Celtiberian Proto-Celtic *\*mātīr* f.n. ‘mother’. Proto-Italic *\*mātēr*, Proto-Slavic *\*màti*, Proto-Celtic *\*mātīr*, and their cognates in Armenian, Hittite, etc. are then traced back to Proto-Indo-European *\*méh<sub>2</sub>-tr-* f.n. ‘mother’. By its completion, this quite revolutionary two-storied and (on the first floor) multi-flat dictionary edifice will serve as a definite replacement of Pokorny’s outdated but still highly valuable *Indogermanisches etymologisches Wörterbuch (IEW)*.

Dictionaries devoted to the inherited lexicon of language families will be able to achieve a high level of excellence if the subgrouping of the cognate languages with which they deal is perfectly established. On the other hand, they are most helpful precisely in establishing these genetic relationships. Thus inheritance dictionaries like *The Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT)*, whose goal is to reconstruct the ancestor language of over 200 languages spoken in South and Southeast Asia whose subgrouping is to the present day controversial, are of particular academic interest, as can be seen in the first part of this dictionary project, *The Tibeto-Burman Reproductive System: Toward an Etymological Thesaurus (STEDTRepr)*, which presents etymologies relating to reproductive anatomy. An earlier publication, the *Handbook*

of *Proto-Tibeto-Burman* (Matisoff 2003), conceived as a sort of companion to the *STEDT* project, received however quite strong criticism because of structural flaws like the lack of explicitness and thus of falsifiability, no safeguards against loans, and faulty Chinese comparisons (see Sagart 2006). In respect to this last issue, the *STEDT* should in any case be consulted in parallel not only with Axel Schuessler's *ABC Etymological Dictionary of Old Chinese (ABCChinese)*, but also with Laurent Sagart's own work *The Roots of Old Chinese* (Sagart 1999). Without being a proper etymological dictionary, this book, which represents a major breakthrough in the field of Chinese etymology, etymologizes hundreds of lexical units pertaining either to the basic vocabulary (personal pronouns, numerals, body parts etc.) or to culturally relevant terms (transportation, commerce, writing etc.).

### **20.2.2. Borrowings**

There is no lack of (more or less etymologically oriented) dictionaries of borrowings, some of them including also loan translations (calques), semantic loans, and loan blends. Be it in loanword dictionaries or in general etymological dictionaries, the lexicographic treatment of borrowings has to pay close attention to dating: in principle –i.e. if the donor language benefits from an as well-documented historical record as the borrowing language–, in order to lend credit to the proposed etymology, the etymon has to be documented before the



loanword. In practice, however, dating borrowings and their etyma is far from being standard practice: only the most sophisticated dictionaries, like the *FEW* and the *LEI*, do it systematically. This is the case for instance in Flöss & Pfister 2012 in *LEI* 12, 1553-1557, CATHEDRA/CATECRA, where Italian CATTEDRALE adj. ‘pertaining to the seat of a bishop’s office’ is dated from the first half of the 14<sup>th</sup> century, and its etymon, Middle Latin CATHEDRALIS, from the 11<sup>th</sup> century; Italian 「SESLONGA」 f.n. ‘reclining chair’, from 1830, and its etymon, French CHAISELONGUE, from 1710. But strictly speaking, the indication of one not contextualized dating for a borrowing is of little significance. First, most datings are tentative and should therefore themselves be dated: each text edition hitting the market contains potentially its allotment of antedatings. If most readers of etymological dictionaries are aware of that, they are probably less mindful of another limitation of datings provided by dictionaries: even if a given dating holds as an absolute starting point, it says nothing about the –often quite lengthy– period between the first time a borrowing was used and its acceptance by the speaking community as a whole. Thus, one cannot but agree with Philip Durkin’s claim that “ideally, etymologies of borrowed items will account for such factors, explaining not only the initial adoption of a word, but its subsequent spread within the lexical system” (Durkin 2009: 163), although very few etymological dictionaries go into such

details. The *Deutsches Fremdwörterbuch (DFWb)*, an etymological dictionary of foreignisms, goes very far in that respect. The entry *Hierarchie* from volume 7 (2010), for instance, which covers twelve pages of text (concerning as well derivatives like HIERARCH, HIERARCHISCH, or HIERARCHISIERN), quotes 26 attestations, from the 13<sup>th</sup> century to 2009, for HIERARCHIE1 ‘angels divided into orders’, 24, from 1533 to 2003, for HIERARCHIE2 ‘ruling body of clergy organized into orders’, and 17, from 1758 to 2009, for HIERARCHIE3 ‘classification of a group of people according to ability or to economic, social, or professional standing’.

Another very nice example is Manfred Höfler’s *Dictionnaire des anglicismes (DictAngl)*, a model in many regards. In this dictionary, three stages of lexicalization are distinguished: quoted lexemes explicitly attributed to foreign languages (marked by square brackets), occasional borrowings in texts (marked by ♦), and borrowings which appear in the word-list of general dictionaries (marked by ||). An example of the first stage can be found in the *DictAngl* s.v. *hurdler*: “[Le *hurdler*, comme les Anglais nomment ce genre de coureur [...]]”, a quotation (‘the *hurdler*, as the English call this kind of runner’) from 1889. As for French PACEMAKER m.n. ‘electrical device for stimulating or steadying the heartbeat’, it is dated as “♦ 1962 *pace maker*; 1964 || Quill. 1965; Rob. S. 1970 *pace maker*”, i.e. the first textual testimony of PACEMAKER dates

from 1962 and is written <pace maker>, whereas the modern spelling <pacemaker> goes back to 1964. The lexicographic acceptance of the borrowing can be dated to the *Dictionnaire encyclopédique Quillet* from 1965 in the modern spelling and to the 1970 *Supplément* of Robert's dictionary in the now disused written form <pace maker>.

Most borrowing processes include more or less extensive phonological and/or morphological accommodation. Ideally, etymological dictionaries would point these out (see Buchi 2006), but at least in print dictionaries, space limitation means this is seldom the case. One exception is provided by the *Dictionnaire des emprunts au russe dans les langues romanes* (*DictEmprRuss*), whose entries are punctuated by tags like “adapt. morph.” (morphological adaptation), “chang. cat.” (change in part of speech), “chang. genr.” (change in gender), “chang. suff.” (suffix change), “greffe suff.” (graft: simplex falsely analyzed as a derivative which received, in place of its pseudo-suffix, a real one).

### **20.2.3. Internal creations**

Within the three grand etymological classes, internal creations receive the least complete etymological coverage: quite often, they are simply listed, without further comment, in a “derivatives and compounds” section under their base (see 20.3.2.). Only etymological dictionaries aimed at specialists apply to internal creations the same scholarly standards as to

inherited lexicon and borrowings. That is the case, for instance, for Gábor Takács's *Etymological Dictionary of Egyptian* (*EDE*), which provides not only explicit etymologies (about base and affix) for the derivatives it contains, but supplies also extensive references to the relevant literature (an advantage perhaps partly explained by the fact that this dictionary is dealing with a chronologically remote language stage, where little can be taken for granted): "derives (by prefix *m-*), as pointed out by H. Grapow (1924, 24), H. Smith (1979, 162), and P. Wilson (PL), from Eg[yp]tian] *nhp* 'bespringen (vom Stier), begatten (vom Menschen)' (O[ld] K[ingdom], Wb II 284, 3-4) = 'to copulate' (FD 135) = 'to procreate' (Smith)" (*EDE* s.v. *mnhp* n. 'procreator'), the only missing information being here the semantic value of the prefix at issue.

Depending on the available sources and their datability, etymological dictionaries may provide first attestations for internal creations, thus enabling the reader, as affixes are only productive during determinate periods, to appreciate the accuracy of the proposed etymologies. In his *FEW* entry of 62 pages devoted to French BALANCE n. 'scales', its cognates and their derivatives and compounds, Jean-Paul Chauveau in *FEW* 2006 s.v. \*BILANX (<http://stella.atilf.fr/few/bilanx.pdf>) thus provides not only explicit etymologies, but also datings (where available, i.e. mostly for French and Occitan) for derivatives,

like BALANCETTE (circa 1180; + -ITTU), BALANCERIE (1415; + -erie), or BALANCIER (1292; + -ĀRIU).

Time depth of etymological dictionaries of languages whose documentation goes back only to recent periods is of course shallower than that of the *FEW*, but this is only a difference of degree and not a difference of kind. For instance, the *Dictionnaire étymologique et historique de la langue des signes française (DEHLSF)* traces back many of the signs of its word-list only to the 18<sup>th</sup> (e.g. ‘connaître’, ‘difficile’, or ‘nuit’) or even to the 19<sup>th</sup> century (e.g. ‘effacer’, ‘fatigué’, or ‘poésie’). If derivatives and compounds are, as a general rule, properly etymologized (i.e. if they are explicitly etymologized!), etymological dictionaries often struggle with less central classes of internal creations like ellipses, clippings, or blends. As for idioms, they often lack completely any etymological analysis, the worst being pragmatemes like English OH, BOY ! interj. ‘(cry of surprise, disappointment, or excitement)’, which is only dealt with in Liberman’s very specialized *Analytic Dictionary of English Etymology (ADEE)* of 55 entries (*ADEE* 17-18). The appearance of new meanings is hardly ever considered worth mentioning (see 20.3.3.).

### **20.3. Current issues in etymography**

In this section, I will discuss a few topics which seem at the same time central for theory and practice of etymological dictionary making and still lacking a conclusive and widely

accepted solution. These thoughts aim to contribute to “the periodic cleansing and, if necessary, the bold replacement of antiquated tools” used by etymographers as advocated by Malkiel (1976: vii). Problematizing these questions at a cross-linguistic level and, ultimately, disregarding possible language related specificities, means that I defend the idea of general etymology (like general phonetics or general semantics) being a viable concept. True, owing to the strong need in this field of work of extensive language-specific knowledge in areas like historical grammar or philological data, etymologists are of necessity permanently attached to a language or at least to a language family. But cross-linguistic collaboration will most certainly yield interesting findings both about general mechanisms of language evolution and about techniques of detecting occurrences of them in order to firmly establish etymologies.

### **20.3.1. Underlying definition of etymology**

The first issue I shall raise is on a very general level and concerns the underlying definition of etymology (see Alinei 1995) shown by etymological dictionaries. Basically, there are two possibilities: etymology can be seen as “that branch of linguistic science which is concerned with determining the **origin** of words” (*OED*)<sup>4</sup> or as “the branch of linguistics which investigates the **origin and history** of words” (*Dictionary of Historical and Comparative Linguistics, DHCL*). According to

the *DLex*, most etymological dictionaries tend to operate on the basis of the second definition: “the emphasis [...] is on the original form of the word (also called its ROOT or ETYMON), but often its whole history or ‘curriculum vitae’ is documented” (*DLex s.v. etymological dictionary*). Indeed, no self-respecting Romance etymologist, for instance, would agree on anything other than a history-oriented definition of etymology. This conception goes back to a paradigm change formalized by Baldinger (1959) and introduced mainly by von Wartburg (through his *FEW* masterpiece) and by Gilliéron, who ridiculed the previous approach to etymology by comparing it to a biography of Balzac consisting of the two following sentences: “Balzac, sitting on his nanny’s knees, was dressed in a blue-and-red striped gown. He wrote *The Human Comedy*” (Gilliéron 1919: 133).

As it is, though, only a very small group of etymological dictionaries –amongst them the *FEW*, the *LEI*, and the *Dictionnaire Étymologique de l’Ancien Français (DEAF)*– practice in a consistent manner “etymology-history of words”, as Baldinger (1959: 239) labelled this, at the time, novel kind of etymology, and practically no one-volume etymological dictionary does, a noteworthy exception being the *OED*-based *Oxford Dictionary of English Etymology (ODEE)*. In this dictionary, indeed, the reader will not only find information, e.g., about the origin of the noun PIRATE (Latin PĪRĀTA), but

also about its semantic enrichment from ‘sea-robber’ in the 15<sup>th</sup> century via ‘marauder’ (16<sup>th</sup> century) to ‘(literary or other) plunderer’ in the 18<sup>th</sup> century.

### **20.3.2. Word-list**

Even today, etymological dictionaries are mostly published on paper, and usually in prestigious (and costly) premium editions. This adds to their respectability and durability, but limits available space, which has direct consequences for the word-list: “etymological information [...] is often omitted from derivatives [...] which are treated as RUN-ON ENTRIES” (*DLex* s.v. *etymological information*). This seems to me very risky, because only a proper etymological analysis can establish that a vocable which presents itself synchronically as a derivative is not inherited or borrowed and represents thus the result of an internal derivation: etymologically speaking, there is no such thing as a transparent derivative! And such a proper etymological analysis will be prevented for vocables which do not appear in the word-list. For that reason I disagree with Malkiel’s assessment (1976: 4) that “furnishing of a separate etymological base for each member of a family, is scientifically unhelpful”: on the contrary, I would plead in favor of granting entry status to all vocables, including derivatives.

Some etymological dictionaries go even further in their groupings. For instance, the *Etymologisches Wörterbuch des Ungarischen* (*EWUng*) presents in the same entry macro-



etymologically linked vocables with distinct etymologies, for instance the Latinism TENOR m.n. ‘voice between bass and alto; tenor singer; tone; content’ and the probable Germanism TENORISTA m.n. ‘tenor singer’ (Gerstner 2002: 572; 579). Such practices should be avoided, be it only because they make automatic extraction and statistical treatment of etymological classes very hard.

### **20.3.3. Etymological (and etymographical) unit**

What constitutes probably the most important progress margin left for etymological dictionaries is closely linked to the fact that even the best etymologists hardly ever give some thought to the question what constitutes the etymological (and etymographical) unit: is it vocables like TABLE (with all its meanings) or lexemes like TABLE1 ‘article of furniture consisting of a flat top and legs’ (see 20.1.)? In my opinion, individual lexemes and not whole vocables are best hypostatized as etymological and etymographical units (see Buchi 2010b and the almost systematic implementation of this principle in the *TLF-Étym* project, e.g. s.v. *gémination*, where one distinguishes a latinism, a germanism, and an internal creation).

If one accepts this approach, one particular etymological category appears as criminally neglected by the whole profession: semantic evolutions.

Each etymological category requires a specific set of information; for semantic evolutions, two of them seem relevant: first, the direct etymon, that is the (possibly no longer existing) lexeme of the same vocable which constitutes the starting point of the semantic evolution, and secondly hints about its coinage, be it by naming a figure of speech like metaphor or metonymy which worked as a universal semantic mechanism or by cross-linguistic comparison. This latter procedure would greatly profit from the “Catalogue of semantic shifts” gathered at the Institute of Linguistics in Moscow (see Zalizniak 2008). Instead of introducing French SAISIR2 ‘to understand’ (since 1694) loosely in an unnumbered paragraph after SAISIR1 ‘to grasp’ (since circa 1100, von Wartburg 1962 in *FEW* 17, 21ab, \*SAZJAN 2), where the semantic link between ‘to understand’ and ‘to grasp’ remains implicit, one could explain the plausibility of such a semantic shift by cross-referencing it to parallels like English TO CATCH, German BEGREIFEN, Italian AFFERRARE or Russian ПОНЯТЬ, which all present the same semantic evolution (see Zalizniak 2008: 228).

#### **20.3.4. *Etimologia proxima vs. etimologia remota***

In theory, most etymologists would probably be in favor of *etimologia proxima*, i.e. of putting forward direct or immediate etymologies. But in practice, etymological dictionaries are full of examples where the *etimologia remota* approach prevails, for instance in Vasmer’s *Russisches etymologisches*

*Wörterbuch (RussEW)*: “über poln[isch] *malować* aus m[ittel]h[och]d[deutsch] *mâlên*” (s.v. *малевать*) or in Cortelazzo and Zolli’s *Dizionario etimologico della lingua italiana (DELI)*: “dal pers[iano] [...], passato in t[ur]c[o] e diffuso in Europa attraverso il fr[ancese] *taffetas*” (s.v. *taffetà*). The etymological discourse is better focused in *The Concise Oxford Dictionary of English Etymology (CODEE)*, which gives the immediate etymology first: “F[rench] *ménage*, earlier *menaige*, *manaige* [, normal development of] [Proto-]Rom[ance] \**mansiōnāticum*, f[ormed on] L[atin] *mansiō*, -*ōn*” (s.v. *ménage*). In my opinion, only “F[rench] *ménage*”, that is the *etimologia prossima* part of the etymology, is relevant. Indeed, the fact that French MÉNAGE is itself inherited has no bearing on its being borrowed by English: had French MÉNAGE been borrowed from another language or created from French material, the borrowing into English would have occurred exactly in the same way.<sup>5</sup> This holds of course even more for the etymology of the Proto-Romance etymon of MÉNAGE, which is definitely irrelevant. So this information is superfluous by virtue of Grice’s maxim of quantity (Grice 1989). But there is more: as the expertise of an etymologist is inevitably less profound in linguistic areas other than those dealt with in the dictionary he compiles, informing the reader about *etimologia remota* constitutes some form of hubris. In the given example, the only defects concern minor inaccuracies which go back to

the –in this case indirect, as the *CODEE* is based on the *ODEE*, which is itself based on the *OED*– source in Romance etymology (probably the *FEW*) or rather to a general flaw of traditional Romance etymology: as the vowel system of Proto-Romance (the proto-language reconstructed from Romance cognates) was based only on timbre and not on quantity, and as Proto-Romance had no equivalent of written Latin <n> before <s> nor <-m> –to say nothing about the fact that in Proto-Romance, stress was phonological– (Buchi and Schweickard 2011: 630-631), “Proto-Rom. *\*mansiōnāticum*” is unsatisfactory by contemporary standards. But the central problem lies in the fact that the energy and the space allotted to *etimologia remota* is then no longer available for *etimologia proxima*: in this case, even if the *etimologia remota* was flawless, it would not make up for the fact that the reader is left in the dark about the question whether the two lexemes mentioned by the *CODEE*, namely MÉNAGE1 ‘housekeeping’ and MÉNAGE2 ‘domestic establishment’, are both borrowed from French or if one of them developed in English (see 20.3.3.). Unfortunately, this kind of lack of balance is very common cross-linguistically, even in the best available etymological dictionaries<sup>6</sup>, and I would like to strongly advocate its replacement by the *etimologia proxima* approach.

### 20.3.5. Degree of formalization

Most (retrospective) etymological dictionaries use only one level of etymological classifiers. For instance, the *RussEW* etymologizes the lexical units it contains by labels like “aus griech[isch]” (демон), “ursl[awisch]” (свет), “Deminutiv” (гуменцо), or “Verstärkung” (хородом ходить). Similarly, the *DELI* will make statements like “comp[osto]” (s.v. *postvocalico*), “da un imit[ativo]” (*badare*), “da [*secento*]” (*secentismo*), “lat[ino]” (*lago*), “lat[ino] parl[at]o” (*pestello*), “loc[uzione] fr[ancese]” (*enfant terrible*), or “v[o]c[e] dotta, lat[ino]” (*ossequio*). Both dictionaries –and they are by no means alone!– also occasionally go discursive, e.g. *RussEW* s.v. полька (“der Tanz ist 1831 in Prag aufgekommen und den unterdrückten Polen zu Ehren benannt”), where the wording leads the reader to think of the noun as a borrowing from Czech, but neither ‘borrowing’ nor ‘Czech’ are made explicit, or *DELI* s.v. *sanseveria*: “chiamata così in onore di Raimundo di Sangro, principe di Sansevero”, where the entry answers the reader’s supposed cultural curiosity, but says nothing about the signifier, the signified, or the syntax of the etymon, nor the language it pertains to, nor its etymological class.

However, authors of etymological dictionaries pertain, in Swiggers’ (1991: 100) wording, to the species of “gardeners” rather than of “moles”, i.e. rather than being “buried in their etymological investigations”, they make it their profession “to

homogenize the grounds and to collect the harvest”.<sup>7</sup> Thus formalization of their etymological discourse plays a major role. I think it would be both more scientific and more helpful for lay readers if etymological dictionaries adopted a two level model, the first level being reserved for the conceptual three-way division among inherited lexicon, borrowings, and internal creations, each of them then being subdivided into more specific categories. Hopefully that would also prevent etymologists from being absorbed by that “quicksand of tiny facts and petty commitments” described by Malkiel (1976: 82). In any case, I agree with his assessment that “a higher level of formalization in linguistics [...] tends to entail more sharply pointed discussion” (Malkiel 1983: 133).

#### **20.3.6. Bringing etymological dictionaries to an end**

It doesn't seem possible to conclude this chapter without addressing the embarrassing question of the publishing rhythm of etymological dictionaries. In fact, there is an important dichotomy that should be added to the phenomenology of etymological dictionaries, namely that between completed ones and uncompleted ones. Unfortunately, indeed, the most advanced and most accomplished representatives of etymological lexicography tend to be almost impossible to terminate in a satisfactory way. This is the case, for instance, for the *LEI*, the first installment of which was published in 1979 and which covers to date letters A, B and parts of C, D,

and E (as well as the beginning of the part devoted to Germanisms). The same holds for the *DEAF*, which goes back to 1974: under Thomas Städtler's leadership, this dictionary was recently split, after having published letters G to K, into two complementary parts: while letters D-F will be compiled in accordance with *DEAF*'s philologically and linguistically outstanding standards, the remaining (approximately 54,000) lemmata from A-C and L-Z will be published in the timesaving form of a rudimentary semantic classification of the Heidelberg file slips.

It appears we etymologists of the early 21<sup>st</sup> century have a collective duty to carry out: going in search of means of successfully completing etymological dictionaries which seem "unfinishable". Of course, online dictionaries with their unlimited possibilities for adding and correcting data go a long way toward addressing this concern. And let's not forget that no (etymological) dictionary was ever completed without a healthy dose of pragmatism!

#### **20.4. Conclusion**

In conclusion, it has to be emphasized that as a whole, (at least European) etymography has reached an excellent standard.

What shortcomings I was led to point out above seem directly related to the fact that even the best educated and most professional "etymologically-minded lexicographer" (Malkiel 1976: 7) is constantly under some cultural pressure to reach out

to the (supposed) needs of the non-specialist by answering (supposedly naive) questions about origin and history of “words”. This of course sidetracks the etymologist from the real goal of presenting in a dictionary, i.e. in a semiformalized form, results from advanced etymological research. I would thus advocate a firm anchoring of etymological lexicographical work in linguistics, i.e. in science (as opposed to culture). In my opinion, this would also have benefits for the general public, as popularization often means reformulating naive questions in order to answer them in a more pertinent way. Many other theoretical and practical issues of etymological lexicography –to quote just a few, selection within the ever-growing available primary data, inclusion or disregard of proper names, or handling of unknown etymologies– could have been discussed in this (too?) short chapter. But the reader might agree with Malkiel (1983: 127), for whom “the ability to control one’s garrulousness has at all times been a major virtue in an etymologist”.

## **20.5. References**

### **20.5.1. Dictionaries**

- ABC Etymological Dictionary of Old Chinese* (2007). Ed. Axel Schuessler. Honolulu: University of Hawaii Press. [= *ABCChinese*]
- An Analytic Dictionary of English Etymology: An Introduction* (2008). Ed. Anatoly Liberman, with the assistance of J.



- Lawrence Mitchell. Minneapolis: University of Minnesota Press. [= *ADEE*]
- Bolchevik, mazout, toundra *et les autres*. *Dictionnaire des emprunts au russe dans les langues romanes. Inventaire – Histoire – Intégration* (2010). Ed. Éva Buchi. Paris: CNRS Éditions. [= *DictEmprRuss*]
- The Concise Oxford Dictionary of English Etymology* (1986). Ed. Terry F. Hoad. Oxford: O.U.P. [= *CODEE*]
- Deutsches Fremdwörterbuch* (2004-<sup>2</sup> [1913-1988<sup>1</sup>]). Ed. Gerhard Strauß et al. Berlin/New York: De Gruyter. [= *DFWb*]
- The Dictionary of Historical and Comparative Linguistics* (2000). Ed. R. L. Trask. Edinburgh: Edinburgh University Press. [= *DHCL*]
- Dictionary of lexicography* (1998). Ed. R. R. K. Hartmann and Gregory James. London/New York: Routledge. [= *DLex*]
- Dictionnaire des anglicismes* (1982). Ed. Manfred Höfler. Paris: Larousse. [= *DictAngl*]
- Dictionnaire Étymologique de l'Ancien Français* (1974-). Ed. Kurt Baldinger et al. Québec/Tübingen/Paris: Presses de l'Université Laval/Niemeyer/Klincksieck. [= *DEAF*]
- Dictionnaire étymologique et historique de la langue des signes française. Origine et évolution de 1200 signes* (2007). Ed. Yves Delaporte. Paris: Éditions du Fox. [= *DEHLSF*]

- Dictionnaire Étymologique Roman* (2008-). Ed. Éva Buchi and Wolfgang Schweickard. Nancy: ATILF:  
<http://www.atilf.fr/DERom>. [= *DÉRom*]
- Dizionario etimologico della lingua italiana* (1999<sup>2</sup> [1979-1988<sup>1</sup>]). Ed. Manlio Cortelazzo and Paolo Zolli. Bologna: Zanichelli. [= *DELI*]
- Etymological Dictionary of Egyptian* (1999-). Ed. Gábor Takács. Leiden/Boston/Köln: Brill. [= *EDE*]
- Etymological Dictionary of Latin and the other Italic Languages* (2008). Ed. Michiel de Vaan. Leiden/Boston: Brill. [= *IEEDLatin*]
- Etymological Dictionary of Proto-Celtic* (2008). Ed. Ranko Matasović. Leiden/Boston: Brill. [= *IEEDCeltic*]
- Etymological Dictionary of the Slavic Inherited Lexicon*. Ed. Rick Derksen. Leiden/Boston: Brill. [= *IEEDSlavic*]
- Etymologisches Wörterbuch des Ungarischen* (1993-1997). Ed. Loránd Benkő. Budapest: Akadémiai Kiadó. [= *EWUng*]
- Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes* (1922-2002). Ed. Walther von Wartburg et al. Bonn/Heidelberg/Leipzig-Berlin/Basel: Klopp/Winter/Teubner/Zbinden. [= *FEW*]
- Indo-European Etymological Dictionaries Online* (2012-). Ed. Alexander Lubotsky. Leiden/Boston: Brill:  
<http://iedo.brillonline.nl/dictionaries> [= *IEDO*].

*Indogermanisches etymologisches Wörterbuch* (1959/1969).

Ed. Julius Pokorny. Bern/Munich: Francke. [= *IEW*].

*Lessico Etimologico Italiano* (1979-). Ed. Max Pfister and

Wolfgang Schweickard. Wiesbaden: Reichert. [= *LEI*]

*The Oxford Dictionary of English Etymology* (1966). Ed. C. T.

Onions, G. W. S. Friedrichsen and R. W. Burchfield.

Oxford: Clarendon. [= *ODEE*]

*The Oxford English Dictionary* (1884-1928). Ed. Sir James A.

H. Murray, Henry Bradley, Sir William A. Craigie and

Charles T. Onions. *Supplement and Bibliography* (1933).

*Supplement* (1972-1986); ed. Robert W. Burchfield. 2nd.

ed., (1989); ed. John A. Simpson and Edmund S. C. Weiner.

*Additions Series*, (1993-7); ed. John A. Simpson, Edmund S.

C. Weiner and Michael Proffitt. 3rd. ed. (in progress) *OED*

*Online* (March 2000-), ed. John A. Simpson, [www.oed.com](http://www.oed.com).

[= *OED*]

*Romanisches Etymologisches Wörterbuch* (1930–1935<sup>3</sup> [1911-

1920<sup>1</sup>]). Ed. Wilhelm Meyer-Lübke. Heidelberg: Winter. [=

*REW*]

*Russisches etymologisches Wörterbuch* (1953-1958). Ed. Max

Vasmer. Heidelberg: Winter. [= *RussEW*]

*The Sino-Tibetan Etymological Dictionary and Thesaurus*

(2011-). Ed. James A. Matisoff. Berkeley: University of

California: <http://stedt.berkeley.edu>. [= *STEDT*]

- The Tibeto-Burman Reproductive System: Toward an Etymological Thesaurus* (2008). Ed. James A. Matisoff. Berkeley/Los Angeles/London: University of California Press. [= *STEDTRepr*]
- TLF-Étym (révision sélective des notices étymologiques du Trésor de la langue française informatisé)* (2005-). Ed. Nadine Steinfeld. Nancy : ATILF: <http://www.atilf.fr/tlf-etym>. [= *TLF-Étym*]
- Trésor de la langue française. Dictionnaire de la langue du XIX<sup>e</sup> et du XX<sup>e</sup> siècle (1789–1960)* (1971-1994). Ed. Paul Imbs and Bernard Quemada. Paris: Éditions du CNRS/Gallimard (see also <http://atilf.atilf.fr/tlf.htm>). [= *TLF*]

### **20.5.2. Secondary literature**

- Alinei, Mario (1995). ‘Thirty-five definitions of etymology or: Etymology revisited’, in W. Winter (ed.), *On Languages and Language. The Presidential Addresses of the 1991 Meeting of the Societas Linguistica Europaea*. Berlin/New York: Mouton de Gruyter: 1-26.
- Baldinger, Kurt (1959). ‘L’*étymologie hier et aujourd’hui*’, *Cahiers de l’association internationale des études françaises* 11: 233-264.
- Bammesberger, Alfred (ed.) (1983). *Das etymologische Wörterbuch. Fragen der Konzeption und Gestaltung*. Regensburg: Pustet.

- Buchi, Éva (2006). 'Wieviel Wortbildung, wieviel Morphologie verträgt die etymologische Forschung? Bemerkungen zur Beschreibung rumänischer Slavismen', in W. Dahmen et al. (ed.), *Lexikalischer Sprachkontakt in Südosteuropa. Romanistisches Kolloquium XII*. Tübingen: Narr: 73-90.
- Buchi, Éva (2010a). 'Where Caesar's Latin does not belong: a comparative grammar based approach to Romance etymology', in Ch. Brewer (ed.), *Selected Proceedings of the Fifth International Conference on Historical Lexicography and Lexicology held at St Anne's College, Oxford, 16-18 June 2010*. Oxford: Oxford University Research Archive (<http://ora.ox.ac.uk/objects/uuid%3A237856e6-a327-448b-898c-cb1860766e59>).
- Buchi, Éva (2010b). 'Comment mesurer le degré d'intégration d'un emprunt linguistique? Une investigation méthodologique sur la base des russismes romans', unpublished paper presented at the international conference «Identifying & Describing Lexical Borrowing» (Liège, March 18-20 2010).
- Buchi, Éva and Schweickard, Wolfgang (2011). 'Ce qui oppose vraiment deux conceptions de l'étymologie romane. Réponse à Alberto Vârvaro et contribution à un débat

- méthodologique en cours’, *Revue de linguistique romane* 75: 628-635.
- Chambon, Jean-Pierre (2010). ‘Pratique étymologique en domaine (gallo-)roman et grammaire comparée-reconstruction. À propos du traitement des mots héréditaires dans le TLF et le FEW’, in I. Choi-Jonin et al. (ed), *Typologie et comparatisme. Hommages offerts à Alain Lemaréchal*. Louvain/Paris/Walpole: Peeters: 61-75.
- Durkin, Philip (2009). *The Oxford Guide to Etymology*. Oxford: Oxford University Press.
- Forssman, Bernhard (1990). ‘Das etymologische Wörterbuch rekonstruierter Sprachen’, in F. J. Hausmann, O. Reichmann, H. E. Wiegand and L. Zgusta (ed.), *Dictionaries. An International Encyclopedia of Lexicography*. Berlin/New York: De Gruyter: 2: 1335-1342.
- Gerstner, Károly (2002). ‘Über die etymologisch-lexikographischen Prinzipien und Methoden des Etymologischen Wörterbuches des Ungarischen’, in A. Braasch and C. Poulsen (eds.), *The Tenth EURALEX International Congress (Copenhagen – Denmark, August 13-17, 2002). Proceedings*. Copenhagen: Center for Sprogteknologi: 1: 569-579.
- Gilliéron, Jules (1919). *La faillite de l’étymologie phonétique: résumé de conférences faites à l’École Pratique des Hautes Études*. Neuveville: Beerstecher.

- Grice, Paul (1989). *Studies in the Way of Words*. Harvard: Harvard University Press.
- Kramer, Johannes (2011). 'Latein, Proto-Romanisch und das DÉRom', *Romanistik in Geschichte und Gegenwart* 17: 195-206.
- Malkiel, Yakov (1976). *Etymological Dictionaries. A Tentative Typology*. Chicago/London: The University of Chicago Press.
- Malkiel, Yakov (1983). 'Models of etymological dictionaries: abandoned, thriving, or worthy of an experiment (exemplified chiefly with Romance)', in Bammesberger 1983, 117-145.
- Malkiel, Yakov (1990). 'Das etymologische Wörterbuch von Informanten- und Korpusssprachen', in F. J. Hausmann, O. Reichmann, H. E. Wiegand and L. Zgusta (ed.), *Dictionaries. An International Encyclopedia of Lexicography*. Berlin/New York: De Gruyter: 2: 1323-1334.
- Matisoff, James A. (2003). *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibeto-Burman reconstruction*. Berkeley: University of California Press.
- Mel'čuk, Igor' A. (2012/2013). *Semantics: From meaning to text*. 2 volumes. Ed. David Beck and Alain Polguère. Amsterdam/Philadelphia: Benjamins.
- Sagart, Laurent (1999). *The Roots of Old Chinese*. Amsterdam/Philadelphia: Benjamins.

- Sagart, Laurent (2006). 'Review of *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibeto-Burman reconstruction*. By James A. Matisoff. Berkeley: University of California Press, 2003. Pp. xlii, 750', *Diachronica* 23: 206-223.
- Schweickard, Wolfgang (ed.) (2011). 'Thematic Part: Historical lexicography of European languages: state of the art and perspectives', *Lexicographica* 27: 1-239.
- Swiggers, Pierre (1991). 'L''étymologie (g)allo-romane: perspectives et points de vue', *Travaux de linguistique* 23: 97-103.
- Thim, Stefan (2011). 'Historical dictionaries of English', in Schweickard 2011, 63-99.
- Van Langendonck, Willy (2007). *Theory and Typology of Proper Names*. Berlin/New-York: De Gruyter Mouton.
- Vàrvaro, Alberto (2011). 'La "rupture épistémologique" del DÉRom. Ancora sul metodo dell'etimologia romanza', *Revue de linguistique romane* 75: 623-627.
- Zalizniak, Anna A. (2008). 'A catalogue of semantic shifts. Towards a typology of semantic derivation', in M. Vanhove (ed.), *From Polysemy to Semantic Change. Towards a typology of lexical semantic associations*. Amsterdam/Philadelphia: Benjamins: 217-232.



---

<sup>1</sup> Many thanks to the very fine lexicographers (and linguists!) who agreed to react to a first draft of this chapter, first of all to Philip Durkin, to whom I am greatly indebted, but also to Jean-Paul Chauveau (Nancy), Steven N. Dworkin (Ann Arbor), Yan Greub (Nancy), Roger Lass (Cape Town), Alain Polguère (Nancy), Laurent Sagart (Paris), Wolfgang Schweickard (Saarbrücken), and Thomas Städtler (Heidelberg).

<sup>2</sup> In case of homonymy, each vocable is numbered separately, e.g. HANGER<sup>1</sup> n. ‘wood on a steep bank’ < Proto-Germanic HANGIAN (*CODEE*) vs. HANGER<sup>2</sup>1 n. ‘one who hangs’ and HANGER<sup>2</sup>2 n. ‘pendent or suspending object’ < English (TO) HANG + -ER (*CODEE*).

<sup>3</sup> To date, ten of them are published: Armenian, Greek, Hittite, Latin, Luvian, Old-Frisian, Proto-Celtic, Proto-Iranian (verbs), Proto-Nostratic, and Slavic.

<sup>4</sup> All boldfaces are mine.

<sup>5</sup> Well-established etymologies lend of course credibility to possible etyma (see Durkin 2009: 170), but that does not necessarily mean they have to be quoted extensively: explicit or even implicit references to the relevant reference works serve the same purpose.

<sup>6</sup> Thim (2011: 90, footnote 31) comes to the same conclusion concerning the *ADEE*: “Although the problem is by no means restricted to them, the Romance borrowings in particular raise the question whether users of a historical dictionary of English need to be given the *etimologia remota* when the immediate source of the borrowing, which after all is the much more relevant information with regard to the history of English, is so often neglected or misrepresented.”

<sup>7</sup> Swiggers (1991: 100): “peut-on parler de types d’étymologistes (personnellement, je vois au moins deux types essentiels: les ‘*taupes*’ enfouies dans leurs recherches étymologiques; les ‘*jardiniers*’ homogénéisant le terrain et rassemblant les récoltes)”.

### **Academic biography**

Éva Buchi is a Senior Researcher at CNRS and a lecturer at Université de Lorraine. She graduated in 1994 at University of Berne with a PhD on the *Französisches Etymologisches Wörterbuch* and in 2003 got a HDR at University of Sorbonne. She specializes in Romance etymology, be it inherited lexicon (*Dictionnaire Étymologique Roman*), borrowings, especially from Slavic languages (*Dictionnaire des emprunts au russe dans les langues romanes*), or internal creations (in particular coining of pragmatemes).

### **Abstract**

This chapter about etymological dictionaries covers mainly two topics. First, it provides, based mostly on examples from European languages, a broad analysis of contemporary practices in etymographical work concerning turn in turn inherited lexicon, borrowings, and internal creations, i.e. the three grand etymological classes which make their own different demands of an etymologist. Then it tackles some issues the author considers of particular interest in current etymology: the dictionary's underlying definition of etymology, the word-list, what should be considered the etymological (and etymographical) unit, *etimologia proxima* vs. *etimologia remota*, the degree of formalization, and the prickly question of bringing etymological dictionaries to an end.

### **Key words**

Borrowings, comparative reconstruction, etymography,  
etymology, inherited lexicon, internal creations, loanwords

### **List of names and subjects**

**Names:** Baldinger (Kurt), Chambon (Jean-Pierre), Chauveau (Jean-Paul), Cortelazzo (Manlio), Durkin (Philip), Gilliéron (Jules), Höfler (Manfred), Leyden University, Liberman (Anatoly), Malkiel (Yakov), Matisoff (James A.), Meyer-Lübke (Wilhelm), Pfister (Max), Pokorny (Julius), Sagart (Laurent), Saussure (Ferdinand de), Schuessler (Axel), Städtler (Thomas), Takács (Gábor), Vasmer (Max), Wartburg (Walther von), Zolli (Paolo)

**Subjects:** borrowings, comparative reconstruction,  
etymography, etymology, general etymology, inherited lexicon,  
internal creations, loanwords

### **Abbreviations**

adj.	adjective
f.n.	feminine noun
interj.	interjection
m.n.	masculine noun
n.	noun
trans.	transitive
v.	verb