Where Caesar's Latin does not belong:
a comparative grammar based approach to Romance etymology
Éva Buchi, ATILF (CNRS & Nancy-Université)

1. Introduction

In this presentation, I would like to argue in favour of a comparative grammar based approach of Romance etymology. In fact, I would not hesitate to assert that what this talk advocates is a paradigm shift in Romance inherited etymology. Let me explain why.

The leading paradigm in etymology of inherited lexicon all over the world is comparative grammar, a classical bottom-up approach, where the common ancestor of a language family is reconstructed from current languages (see Fox 1995). Indeed, the comparative method is successfully applied in Indo-European, Germanic, Slavic, Semitic, Austronesian, Sino-Tibetan… etymology.

But what is true for almost all language families does not apply to Romance languages. In this domain scholars discard the comparative method as unnecessary in the face of all the written testimonies of classical Latin, from Plautus via Caesar to Tacitus. Instead, they apply a top-down method, which stresses the disintegration of 'high' Latin into 'low' Romance languages. So, since its beginning in the 19th century, when dealing with inherited lexicon, Romance etymology and etymography always promoted classical Latin etyma (like *caballus*, *dĕcĕm* or *hĕrba* [REW₃]) and gave credit to the comparative method only in those rare occasions when there is no handy Latin etymon available (like *abbĭbĕrāre* [REW₃]). So, while comparative reconstruction is a well-established branch of linguistics with well-tried principles and tools, it is in general not applied to Romance etymology.

However, in two quite revolutionary papers, Jean-Pierre Chambon (2007; 2010), a professor at the Sorbonne, recommended recently recreating Romance etymology on the basis of comparative grammar. The purpose of this paper is to validate Chambon's claim that, when dealing with inherited lexicon, Romance etymologists should use the comparative method. I will do that by applying comparative grammar to one set of Romance cognates.

2. Etymology of Italian *cadere* 'to fall' and its cognates

If you are interested in the etymology of the English verb *(to) fall*, you may look it up in the OED₂ (or in OED online) or, if you want the short version, you may open the *Oxford Dictionary of English Etymology* (Onions *et al.* 1966). You will see the authors of this dictionary first enumerate several cognates of *(to) fall*: Old Frisian and Old Norse *falla*, Old Saxon and Old High German *fallan*. They then conclude that this cognate series shares the etymon *fallan*, which is a Common German or Proto-German lexeme. The fact that this common ancestor is reconstructed is marked by an asterisk.

Let us have a look at the etymology provided for the semantic equivalents of *(to) fall* in Romance languages: Italian *cadere*, French *choir*, Spanish *caer* and so forth. The reference dictionary of Romance etymology, Meyer-Lübke's almost hundred-year-old *Romanisches Etymologisches Wörterbuch* (REW₃), gives the following information:

> 1451. **cadĕre** 'fallen', 2. ***cadēre**.
> 2. Rum. *cădeà*, vegl. *kadar*, it. *cadere*, friaul. *kadér*, frz. *cheoir*, prov. *cazér*, kat, *cáurer*, sp., pg. *caer*, npg. *cahir*; rum. *se cade* 'es schickt sich', vgl. *61*. Das Wort ist im Lomb.,

> Venez. und Rätorom. selten, in Frankreich durch *tomber*, *tumbá* mehr und mehr an die Peripherie gedrängt. – Ablt.: […]

As you will see, in Romance etymology, the comparative method is bypassed by mentions of Latin etyma. In the REW₃, the headword under which Italian *cadere*, French *choir* and Spanish *caer* are treated is *cadĕre*, a classical Latin verb which can be found in any Latin dictionary. But then the REW mentions a second etymon, the subentry *cadēre, which is some sort of 'fiddled with' classical Latin: the short <e> has been replaced by a long one in order to accommodate the Romance forms. This is the closest traditional Romance etymology comes to applying comparative reconstruction: as a means to fill gaps in cases where textual tradition is lacking exact Latin correlates. The underlying idea is however clearly that the etymon pertains to Latin: graphic presentation is given according to Latin conventions (the velar plosive /k/ is noted by the grapheme <c>) and, more to the point, the etymon comprises a long /e/, which is consistant with the Latin vowel system, whereas everybody agrees that having long and short vowels is not a feature presented by the common ancestor of the Romance languages (commonly named 'Vulgar Latin').

Note that, as all the Romance lexemes are filed under the subentry 2., the main headword does not account for Italian *cadere* nor for any of its cognate forms: one has every right to question the legitimacy of that headword.

By the way, in the domain of Romance linguistics the asterisk usually (as in the REW₃) does not mean 'reconstructed', but 'not attested': written testimony –or, as in this case, the absence of written testimony– is paramount.

My claim is that applying the comparative method to this etymological family deepens our understanding of its origin. I will limit the demonstration to the stressed syllable.

Italian (Tuscan), Venetian and Sicilian [ka'dere] and old Catalan [ka'der] present the stressed syllable /'de/. Romanian [kə'dea] shows diphtongization and Dalmatian [ka'dar] opening to /a/. Old Occitan [ka'zer] presents the fricative /z/ instead of the plosive /d/, and in Spanish and Galician-Portuguese [ka'er], even the fricative dropped. As to French /wa/, it is a regular (if quite exotic) evolution, like in *croire* as opposed to Italian *credere* and Spanish *creer*. All these stressed syllables stem, as they represent regular sound correspondences, from Proto-Romance (the reconstructed antecedent language) */'de/, and by applying the same method to the whole lexeme, we get */ka'd-e-re/.

According to the REW₃, */ka'd-e-re/ is the only etymon to invoke. Yet this is a quite incomplete view. Indeed, Venetian ['kaze], Sicilian ['kadiri], Occitan ['kaire] and Catalan ['kaurə] represent variants with stressed syllables /ka/. These forms stem of necessity from Proto-Romance */'kad-e-re/. Indeed, morphologic analysis shows that we are in presence of two inflectional types (*/'kad-e-re/ and */ka'd-e-re/) of one single verb, which I propose to present in the more abstract form of Proto-Romance */'kad-e-/. If we consider both of these inflectional types, it appears that with the exception of Sardinian, Romanche and Ladin, all Romance idioms present heirs of this Proto-Romance */'kad-e-/ '(to) descend freely by the force of gravity, (to) fall'.

Geolinguistics shows that the */'kad-e-re/ type, which is only present in the geographical centre of the Romance area, is recessive. It pertains to an older stratum, whilst the */ka'd-e-re/ type, which is ubiquitous, translates in a younger stratum.

I will now correlate this linguistically motivated stratigraphy with two independent sets of data. The first of them is extra-linguistic, since it pertains to geographical areas conquered by the Romans at different periods. Consider the map provided by Manfred Raupach in the *Lexikon der Romanistischen Linguistik* (Raupach 1996), which represents the latinization process in the antiquity: in 300 before Jesus Christ, Latin was only spoken in Rome and in the Latium region; after a hundred years of expanding, Latin had spread to the southern part of the Italic peninsula and to Sicily; in 100 before Jesus Christ, Latin was spoken as well in the northern part of the Italic peninsula, in the South of Gallia and in the Iberic peninsula; only around Jesus Christ, Latin spread to the North of Gallia and to the future homelands of Dalmatian and Romanian.

The */'kad-e-re/ type is found only in anciently latinized areas, whereas the */ka'd-e-re/ type is everywhere (or was, at least, in the Middle Ages). That shows that the most ancient variety of Proto-Romance knew both inflectional types, while in more recent Proto-Romance the verb only lived in the inflectional variant */ka'd-e-re/.

This finding relies on analyses made by Romance linguistics and by historians of the Roman empire. Let's now draw in Latin philology. Written Latin data confirms this chronology. The correlate of */'kad-e-re/, *cadere*, is known throughout all of antiquity: it is attested since Ennius (TLL 3, 16-32). As to the second inflectional type, its correlate, *cadēre*, is attested only in the late antiquity, in the 4th century A.D. (*cadebit* [4th century]; *cadeat* [ca 400]; TLL 3, 16; Stotz 1998: 186). This is consistent with the hypothesis that */'kad-e-re/ and */ka'd-e-re/ represented variants within the Latin diasystem, where */'kad-e-re/ belonged to the acrolect and */ka'd-e-re/ to the basilect. One can assume that in the earlier periods, the Roman conquerors disseminated a form of spoken Latin which contained both variants, whereas in the more recent periods, when expansion of the Roman empire relied less and less on native speakers and more and more on bilinguals whitout access to the acrolect, only */ka'd-e-re/ survived.

3. Evaluation of the efficiency of the comparative method

The purpose of our study was to validate Chambon's claim that, when dealing with inherited lexicon, Romance etymologists should use the comparative method. So was Chambon right in pleading for the replacement of the traditional 'look it up in the Latin dictionary' approach of Romance etymology by the methodological framework of linguistic reconstruction? And in treating Romance languages like 'normal' languages?

Our findings concerning the etymology of Italian *cadere* and its cognates clearly show that yes, the comparative method works just fine for Romance etymology, and it yields more interesting results than the Latin-centered and grapho-centered method practiced until now.

First for Romance etymology: this method enables us to establish not only the existence of a lexeme made of two inflectional types, which differ in their belonging to two distinct diastratic varieties of Proto-Romance, but also their stratification: a more acrolectal one (*/'kad-e-re/), which fades from use as time goes by, and a more basilectal one (*/ka'd-e-re/), which remains common even in later Proto-Romance. As all languages, including reconstructed ones, know internal variation, this finding is anything but mysterious, but only comparative etymology could bring it to the surface (see Buchi & Schweickard to appear). As Philip Durkin put it:

> Even though our surviving records for classical Latin are mostly literary and reflect a highly homogeneous literary language, there is indeed some variation in our surviving Latin evidence, and the later evidence of the Romance languages suggests the existence of a good deal of further variation in Latin which is not reflected in the surviving documentary evidence. (Durkin 2009: 10-11)

But applying the comparative method to Romance languages also yields more interesting results for Latin etymology: instead of just giving Latinists back what they already know, we give them a deeper understanding of their data by placing them in the context of a diasystem.

Finally, this method yields also more interesting results for Indo-European etymology, because Proto-Romance data compare, more easily than Latin data do, to Proto-x data. Together with Proto-Germanic, Proto-Slavic, Proto-Indo-Iranian, and so on, Proto-Romance helps to reconstruct Proto-Indo-European.

4. DÉRom project

So in my eyes comparative grammar is a valuable method to be applied in Romance inherited etymology; I would even say the most valuable. However, you may be tempted to challenge this generalization on the grounds that one single example is not quite that convincing in itself. Indeed, I readily acknowledge that in this study, the comparative method was only applied to one set of Romance cognates. To completely validate the new method, the same procedure should be replicated with other lexemes. This is where a European research programme comes into play as whose ambassador I speak to you today. It is called *Dictionnaire Étymologique Roman*, DÉRom in short (see Buchi 2010; Buchi, Chauveau, Gouvert & Greub 2010; Buchi & Schweickard 2008; 2009; 2010; Florescu 2009; Schweickard 2010): a new Romance etymological dictionary which goes by the comparative method. The goal of DÉRom, during its first stage, consists in reconstructing some 500 etymological bases common to all or nearly all Romance languages: the core lexicon of Proto-Romance. It is edited by Wolfgang Schweickard, a professor of Romance linguistics at Saarland University, and myself, and compiled by a team of about 40 linguists based in eleven countries (Austria, Croatia, Czechia, France, Germany, Italy, Portugal, Romania, Slovenia, Spain and Switzerland). The dictionary is published online, on a free Web site hosted by the ATILF (see DÉRom in the bibliography). For the time being, 32 entries are available, all of which lend support to the idea that comparative grammar is a useful method in the domain of Romance etymology.

Let's just quote some of our more important results:

– Buchi 2009–2010 *in* DÉRom *s.v.* */ˈɸak-e-/ '(to) do' distinguishes an original type */ˈɸak-e-re/ and a syncopated one */ˈɸ-a-re/, the development of which is put in relation with the high frequency of this verb as an auxiliary in proclitic position in the Romance synthetic future */ɸ-a-re-ˈaβ-e-/.

– Medori 2008–2010 *in* DÉRom *s.v.* /ˈkarpın-u/ 'hornbeam' subdivides in two sections, each of which is devoted to an evolved form of the original feminine noun */ˈkarpın-u/: first the masculine noun */ˈkarpın-u/, which is the result of integrating the lexeme in a more prototypical inflectional class, then the remorphologized feminine noun */ˈkarpın-a/.

– Florescu 2010 *in* DÉRom *s.v.* /ˈlaks-a-/ '(to) let' shows that from the original form */ˈlaks-a-/ stemmed, still in Proto-Romance, a more colloquial form */ˈlaks-i-a-/, which is fitted up with the interfix /-i-/ (see de Dardel 2006).

– Andronache 2008-2010 *in* DÉRom *s.v.* */ˈpɔnt-e/: this entry distinguishes the original masculine noun, a more recent regularized feminine noun, finally a restored masculine, whose emergence explains by acrolectal pressure.

5. Conclusion

In conclusion, I hope you agree that applying Chambon's methodological plea to Italian *cadere* and its cognates bore fruit. In any case, the */ˈkad-e-/ example provided me with the opportunity to present the DÉRom project. At the same time, this paper constituted the occasion for explaining ourselves, for, as Anatoly Liberman put it: 'In conformity with their genre, etymological dictionaries emphasize the results rather than the process of the investigation' (Liberman 2009: 96).

One might think that in Romance etymology, as it is 'easier' than, say, Bantu or even Germanic etymology, there are no major discoveries left to be made. But, as Georgia Green and Jerry Morgan put it in their *Practical guide to syntactic analysis*:

> Beginning students are sometimes discouraged by the belief that 'all the easy stuff's already been done. What's left is really hard.' But when that 'easy stuff' is examined closely, it often turns out that it is only half-done, and that the conclusions do not follow from the premises (which often are not made explicit), or that the assumptions they are based on are no longer considered tenable. A surprising amout of the 'easy stuff' needs to be re-done. (Green & Morgan 1996: 17)

In the same way, our example shows that applying the comparative method to Romance etymology is likely to provide new insight into this branch of linguistics: apparently, because of its Latin-orientedness, Romance inherited etymology was only half-done. This explains, I think, Günter Holtus's and Fernando Sánchez Miret's opinion, as voiced in the following quotation, even if they do not insist particularly on comparative grammar:

> In short, historical-comparative research of the Romance lexicon still stimulates work in Romance linguistics and, as shown by the DÉRom project, is able to offer new and crucial insights in the history of the Romance languages. (Holtus & Sánchez Miret 2008: 178)[1]

In any case, Romance inherited etymology is currently undergoing a major change of paradigm. Using from now on the technique of comparative reconstruction, a method applied to language families all over the world, Romance historical linguistics will definitely be better integrated in general historical linguistics, which will no doubt be highly profitable for both parties.

---

[1] The translation is ours. Original text: "En definitiva, la investigación histórico-comparativa del léxico de las lenguas románicas no ha cesado de estimular el trabajo de los romanistas y, tal como demuestra el projecto del *DÉRom*, está en disposición de ofrecer nuevos frutos esenciales para la comprensión del desarrollo histórico de la Romania".

6. Bibliography

Buchi, Éva (2010): 'Pourquoi la linguistique romane n'est pas soluble en linguistiques idioromanes. Le témoignage du *Dictionnaire Étymologique Roman* (DÉRom)'. In: Alén Garabato, Carmen *et al.* (eds.): *Quelle linguistique romane au XXIe siècle ?* Paris: L'Harmattan: 43-60.

Buchi, Éva, Chauveau, Jean-Paul, Gouvert, Xavier & Greub, Yan (2010): 'Quand la linguistique française ne saurait que se faire romane: du neuf dans le traitement étymologique du lexique héréditaire'. In: Neveu, Franck *et al.* (edd.): *Congrès Mondial de Linguistique Française – CMLF 2010.* Paris: Institut de Linguistique Française (http://dx.doi.org/10.1051/cmlf/2010025): 111-123.

Buchi, Éva & Schweickard, Wolfgang (2008): 'Le *Dictionnaire Étymologique Roman* (DÉRom): en guise de faire-part de naissance'. *Lexicographica. International Annual for Lexicography* 24: 351-357.

Buchi, Éva & Schweickard, Wolfgang (2009): 'Romanistique et étymologie du fonds lexical héréditaire: du REW au DÉRom (*Dictionnaire Étymologique Roman*)'. In: Alén Garabato, Carmen *et al.* (eds.): *La Romanistique dans tous ses états*. Paris: L'Harmattan: 97-110.

Buchi, Éva & Schweickard, Wolfgang (2010): 'À la recherche du protoroman: objectifs et méthodes du futur *Dictionnaire Étymologique Roman* (DÉRom)'. In: Iliescu, Maria, Siller-Runggaldier, Heidi & Danler, Paul (eds.): *Actes du XXVe Congrès International de Linguistique et de Philologie Romanes (Innsbruck, 3-8 septembre 2007).* Berlin/New York: de Gruyter: 6: 61-68.

Buchi, Éva & Schweickard, Wolfgang (to appear): 'Per un'etimologia romanza saldamente ancorata alla linguistica variazionale: riflessioni fondate sull'esperienza del DÉRom (*Dictionnaire Étymologique Roman*)'. In: Boutier, Marie-Guy, Hadermann, Pascale & Van Acker, Marieke (eds.): *Variation et changement en langue et en discours*.

Chambon, Jean-Pierre (2007): 'Remarques sur la grammaire comparée-reconstruction en linguistique romane (situation, perspectives)'. *Mémoires de la Société de linguistique de Paris* 15: 57-72.

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: Choi-Jonin, Injoo, Duval, Marc & Soutet, Olivier (eds): *Typologie et comparatisme. Hommages offerts à Alain Lemaréchal*. Louvain/Paris/Walpole: Peeters: 61-75.

Dardel, Robert de (2006): 'Les variantes lexématiques avec l'interfixe /I/ en protoroman'. *Revue de linguistique romane* 70: 377-407.

DÉRom = Buchi, Éva & Schweickard, Wolfgang (eds.) (2008–): *Dictionnaire Étymologique Roman (DÉRom)*. Nancy: ATILF (http://www.atilf.fr/DERom).

Durkin, Philip (2009): *The Oxford Guide to Etymology*. Oxford: Oxford University Press.

Florescu, Cristina (2009): 'Limba română în *Dictionnaire Étymologique Roman* DÉRom (< *Romanisches Etymologisches Wörterbuch* REW)'. In: Botoşineanu, Luminiţa *et al.* (eds.): *Distorsionări în comunicarea lingvistică, literară şi etnofolclorică românească şi contextul european*. Iaşi: ALFA/Asociaţia Culturală "A. Philippide": 153-159.

Fox, Anthony (1995): *Linguistic Reconstruction. An Introduction to Theory and Method*. Oxford: Oxford University Press.

Green, Georgia M. & Morgan, Jerry L. (1996): *Practical guide to syntactic analysis*. Stanford: Center for the Study of Language and Information.

Holtus, Günter & Sánchez Miret, Fernando (2008): *«Romanitas», Filología Románica, Romanística*. Tübingen: Niemeyer.

Liberman, Anatoly (2009) : 'Dwarves, Pimps, and Galoots, or Chance, Luck, and Serendipity in an Etymologist's Work'. In: McConchie, R. W., Honkapohja, Alpo & Tyrkkö, Jukka (eds.): *Selected Proceedings of the 2008 Symposium on New Approaches in English Historical Lexis (HEL-LEX 2)*. Somerville, MA: Cascadilla Proceedings Project: 96-103.

OED₂ = Simpson, J. A. & Weiner, E.S.C. (eds.) (1989^2 [1933^1]/1993–1997): *The Oxford English Dictionary* (20 volumes and 3 additional volumes). Oxford: Clarendon.

Onions, C. T., Friedrichsen, G. W. S. & Burchfield, R. W. (1966): *The Oxford Dictionary of English Etymology*. Oxford: Oxford University Press.

Raupach, Manfred (1996): 'Expansion und Rückzug des Lateins'. In: Holtus, Günter *et al.* (eds.): *Lexikon der Romanistischen Linguistik (LRL)*. Tübingen: Niemeyer: 2/1: 5-19.

REW₃ = Meyer-Lübke, Wilhelm ($1930–1935^3$ [$1911–1920^1$]): *Romanisches Etymologisches Wörterbuch*. Heidelberg: Winter.

SCHWEICKARD, Wolfgang (2010): 'Die Arbeitsgrundlagen der romanischen etymologischen Forschung : vom REW zum DÉRom'. *Romanistik in Geschichte und Gegenwart* 16: 3-13.

Stotz, Peter (1996–2004): *Handbuch zur lateinischen Sprache des Mittelalters* (5 volumes). Munich: Beck.

TLL = 1900–. *Thesaurus Linguae Latinae*. Leipzig: Teubner.