

VARIA

GILLES QUENTEL

gillesquentel@free.fr

Université de Gdańsk

LES PROBLÈMES MÉTHODOLOGIQUES DE L'ÉTYMOLOGIE D'ORIGINE GAULOISE EN FRANÇAIS

Abstract. Gilles Quentel, *Les problèmes méthodologiques de l'étymologie d'origine gauloise en français* [Methodological problems of the etymology of Gaulish origin in French], *Studia Romanica Posnaniensia*, Adam Mickiewicz University Press, Poznań, vol. XLI/3: 2014, pp. 133-147. ISBN 978-83-232-2759-5. ISSN 0137-2475. eISSN 2084-4158. DOI: 10.7169/strop2014.413.010

It is a well-known fact that the etymology of the French lexicon is mostly Latin. But this apparently obvious consideration hides a significant amount of unsure, if not biased, etymologies. This paper aims first at criticizing the laxity of the etymological method : use of reconstructed vulgar Latin and old Frankish etymons upon unclear principles, huge distortions of meaning, phonetic irregularities, ignorance of the historical context, etc. In a second time, this paper aims at showing that the neglected Gaulish substratum can sometimes afford better explanations to some of these doubtful and impressionistic etymologies, and proposes new etymologies based upon attested Gaulish etymons.

Keywords: historical linguistics, etymology, French, Gaulish, vulgar Latin, Old Frankish

La genèse du français, comme celle des langues romanes en général, est un problème éminemment complexe dans la mesure où elle met en jeu plusieurs langues de substrat et de superstrats autour d'une langue lexicalisante et grammaticalisante, le latin, avec laquelle on ne peut l'articuler sans concevoir l'existence d'une langue transitoire théorique, le latin populaire.

Si le latin est richement documenté, il n'en va pas de même pour les autres : notre connaissance du gaulois est très fragmentaire, le francique nécessite une complète reconstruction sur la base de ses parentes germaniques, et rien n'atteste de l'existence même du latin populaire. A ceci, il convient d'ajouter l'influence probable de substrats pré-celtiques dont on n'a plus de traces que dans la toponymie et dans l'unique survivante de celles-ci : le basque, dont l'impact sur la genèse du gallo-roman est mal élucidé. Dans ces conditions, l'étymologiste doit opérer par conjectures et évaluer la fiabilité de celles-ci au prix d'un subtil exercice d'équilibriste.

1. LES MÉCANISMES FONDAMENTAUX DE LA RECHERCHE ÉTYMOLOGIQUE GAULOISE

1.1. ÉTAT DE L'ART

La quantité de mots gaulois admis par les dictionnaires est fluctuante, et la recherche étymologique en évolution constante. Les notices du TLF donnent un bon aperçu de la méthode employée pour détecter les étymologies gauloises (dont il recense 272 racines), et c'est sans doute la source la plus détaillée que l'on puisse trouver sur le sujet.

La Langue Gauloise de Pierre-Yves Lambert (2003) qui vient remettre à jour la *Langue Gauloise* de Georges Dottin (1918) et le *Dictionnaire de la Langue Gauloise* de Xavier Delamarre (2003) proposent un compte-rendu exhaustif de l'état de l'art en la matière ainsi que de nouvelles étymologies mises en lumière par les récentes découvertes archéologiques (Plat de Lezoux, Tablette de Chamalières, Plomb du Larzac et Tuile de Châteaubleau). La langue des francs n'a pas, jusqu'ici, fait l'objet d'un ouvrage qui lui soit dédié et n'est traitée ponctuellement qu'à travers le prisme de l'étymologie.

1.2. À QUELLES CONDITIONS UN ÉTYMON EST-IL GAULOIS ?

La pertinence d'un étymon gaulois repose sur les cinq conditions suivantes, que l'on peut considérer comme cardinales dans la mesure où la dérogation à l'un d'entre eux écarte *a priori* l'éventualité d'une étymologie gauloise. On verra plus tard dans quelle mesure ces mécanismes peuvent ne pas être toujours fiables, et quels types d'interférences peuvent en invalider les conclusions.

– le mot ne provient pas du latin : c'est en général d'abord dans le latin que l'on cherche la source d'un mot français. Cette recherche est théoriquement assez sûre car le latin possède un vaste corpus littéraire et épigraphique attesté et son évolution phonétique est bien élucidée (Bourcier, 1930 ; Zink, 1986 ; Bonnard, 1992 ; De la Chaussée, 1974).

– le mot est attesté uniquement dans la zone des parlars dits « gallo-romans » (langues romanes de l'ouest européen à l'exclusion du sarde (*logudoro*, cf. Diez 1874 : 88), c'est-à-dire formées sur un substrat celtique (français, provençal, catalan et portugais) ou partiellement celtique (espagnol et italien). Si des cognats du mot recherché existent en roumain ou en sarde, langues apparues sur des substrats respectivement dace (Mihaila, 1978) et ibérique (Blasco Ferrer, 2010), et si ces cognats ne sont pas des emprunts aux langues de type gallo-roman, alors l'hypothèse gauloise est disqualifiée.

– le mot n'a pas de cognat dans une langue de superstrat : langues germaniques et arabe notamment.

– le mot possède un cognat attesté en gaulois ou à défaut dans plusieurs langues celtiques insulaires (brittoniques et/ou gaéliques) permettant sa reconstruction (Lane, 1933). Le mot et ses cognats doivent présenter des divergences phonétiques conformes aux lois de la phonétique des langues indo-européennes afin d'écarter l'hypothèse d'un emprunt ancien au latin ou au germanique.

– le mot procède régulièrement de la forme gauloise. Cette dernière condition est difficile à remplir en raison de la multiplication des exceptions dans la phonétique historique du français d'une part, de l'incertitude quant à la place de l'accent dans les mots gaulois (Lambert, 2003 : 48), de la faible quantité de documents écrits en gaulois et de la connaissance imparfaite de la langue qui en résulte ainsi que de la forme parfois très proche des étymons latins et gaulois.

On pourrait ajouter une cinquième condition : la présence attestée du mot étudié dans une autre langue celtique continentale (lépontique et celtibère), mais cette condition est difficilement réalisable compte tenu de la très faible taille des corpus lexicaux disponibles dans ces langues (Evans & Eska, 1993 ; Lejeune, 1971 ; Untermann, 1997) et des incertitudes qui subsistent quant à leur interprétation.

2. LES INTERFÉRENCES SUBSTRATIQUES ET SUPERSTRATIQUES

Les conditions énoncées plus haut se heurtent, dans la pratique, à un certain nombre de problèmes que nous nous proposons d'examiner plus en détail. Le premier est celui des interférences entre substrats et superstrats. Il est généralement dû à l'ordre de priorité que l'on établit, de façon parfois arbitraire, entre ces différentes strates de formation de la langue française, et qui posent chacune des problèmes spécifiques dans le cadre de la recherche étymologique. Nous allons examiner ces strates et étudier de manière plus précise quelques cas emblématiques.

2.1. LE LATIN

Le latin est la langue-mère du français, il ne s'agit donc pas d'un substrat et la priorité qui lui est octroyée dans la recherche est amplement justifiée. Il existe cependant un certain nombre de mots pour lesquels l'étymon latin proposé est très proche d'un mot gaulois attesté. On en trouve 11 cas dans le Dictionnaire de Xavier Delamarre. Aucun de ces mots gaulois ne semble avoir été emprunté au latin et sauf mention du contraire, leur évolution phonétique vers le français est régulière. La liste ci-dessous en propose les plus significatifs (les étymons PIE sont de Pokorny, 1966) :

Lat. *ad* / Gaul. *ad* [même sens] pour le français « à ». PIE **ad*-

Lat. *alba* / Gaul. *albos* [même sens] pour le français « aube ». PIE **al*-

Lat. *arduus* [élevé] / Gaul. *ardu* [haut] pour le français « ardu ».

Lat. *camurum* [courbé] / Gaul. *cambo* [courbe] pour le français « cambré ». PIE **kam-*.

Lat. *canna* [roseau] / Gaul. *cano* [même sens] pour le français « canne ». [NB : les nasales géminées ne sont pas toujours transcrites en gaulois (Lambert, 2003 : 45)]

*Lat. *centum* / Gaul. *canto* [même sens] pour le français « cent ». PIE **kmtom*.

*Lat. *carus* / Gaul. *caros* [même sens] pour le français « char » et ses dérivés.

*Lat. *gemellus* / Gaul. *iemo*. [même sens] pour le français « gémeau ». PIE **g'em*. Les deux formes sont régulières mais la graphie « -eau » fait peser la balance dans le sens de *gemellus*.

Lat. *juvenis* / Gaul. *iouincos* (qui pose le problème du [k]). PIE **ieu-*.

Lat. *rota* / Gaul. *roto*. Ce mot non-IE serait celtique selon Alinei (2001).

Lat. *vir* / Gaul. *uiros*. PIE **uiro-s*.

Dans certains cas, il est assez aisé de distinguer l'étymon réel grâce à la phonétique. Ainsi dans la paire Lat. *mare* / Gaul. *mori*, l'aboutissement du [o] gaulois au [e] français est impossible. À l'inverse le [a] latin peut sans difficulté passer au [e] français (*mare* / *mer*). Le même scénario se retrouve entre autres dans les paires *rubeus* / *roudo*, *nux* / *cnuu* et *decem* / *decam*. Dans d'autres cas, la présence d'un cognat dans une langue de substrat non-celtique invalide l'étymologie gauloise, comme dans les paires *cervus* / *caruos*, *deus* / *deuos*, *in* / *eni*, *inter* / *enter*, *novus* / *nouuo*, *noctem* / *noxt*, *pretium* / *pritom* et *sacer* / *sacro*.

Pour les autres paires, celles qui figurent dans notre liste, plusieurs scénarios sont envisageables : le plus communément admis est celui de deux mots, l'un latin et l'autre gaulois, ayant évolué parallèlement à partir d'un unique étymon (PIE ou autre), et dont l'aboutissement latin est à son tour l'étymon du français, le gaulois n'ayant alors été oublié.

On peut également défendre l'hypothèse inverse (l'aboutissement gaulois est l'étymon du français) à la condition que le mot ne se trouve pas dans une langue de substrat non-celtique. C'est le cas de *ad*, *ardu*, *cano*, *caros*, *iouincos* et *uiros*. Cette hypothèse reste valide si le mot correspondant en roumain ou en sarde est un emprunt (ro. « *cambra* » emprunté au fr.).

Si le mot existe en roumain ou en sarde, qu'il ne s'agit pas d'un emprunt tardif à une autre langue romane, alors on peut encore imaginer une troisième possibilité : un emprunt du latin au celtique qui peut se justifier pour des raisons civilisationnelles. C'est peut-être le cas pour *argentum* / *argenton* (ro. « *argint* »), *jugum* / *iugo* (ro. « *jug* »), *rota* / *roto* (ro. *roată*) (voir à ce propos Alinei, 2001).

2.2. LE VIEUX FRANCIQUE

Si l'on excepte quelques mots épars dans une *Lex Salica* de 510 et une inscription runique aux origines discutées, il n'existe aucune trace de la langue des anciens Francs. Ajoutons que les études scientifiques dédiées à la question du vieux francique *stricto sensu* sont rarissimes, ce qui pose d'emblée un problème méthodologique. Cette langue n'est envisagée que de façon connexe dans des études à caractère étymo-

logique, et surtout dans les dictionnaires qui tentent d'en reconstruire le lexique par le biais de la phonétique historique.

Que désigne-t-on alors par l'étiquette « vieux francique » ? Il s'agit d'une langue germanique parlée entre le IV^{ème} et le VII^{ème} siècle, probablement originaire de la basse vallée du Rhin, dont l'idiome connu le plus proche est le vieux néerlandais, lequel fut parlé quatre siècles plus tard et attesté pour l'essentiel dans les *Psaumes de Wachtendonck*. Lefait de savoir si le vieux néerlandais est un descendant direct du vieux francique n'est cependant pas clairement établi, quoique certains éléments le laissent penser (notamment la présence d'un cas de dévoisement final dans l'inscription runique de Bergakker ; Bammesberger, 1992 ; Mees, 2002). Cependant, ce n'est pas le vieux néerlandais, peu attesté, qui sert de base à la reconstruction du vieux francique, mais le vieux haut-allemand (*vha*) dans la majorité des cas et le vieil islandais dans la plupart des autres.

Il ressort de tout ceci que 1) l'étymon francique n'est jamais attesté et que 2) il est reconstruit sur la base du *vha* alors que le francique est de toute évidence une forme de vieux bas-allemand (*vba*). Phonétiquement cela ne pose pas de problème : les lois d'évolution phonétique du germanique sont bien documentées et il est en effet possible de reconstruire un mot du *vba* à partir *vha*, en tenant notamment compte du fait que la seconde mutation consonantique n'a pas eu lieu en *vba*. Mais le mot reconstruit sera une sorte de modèle mathématique, car lexicalement, rien *a priori* ne dit qu'un mot du *vha* a existé sous la même forme en vieux francique : les deux langues n'ont pas les mêmes substrats ni opéré les mêmes emprunts et n'ont pas forcément suivi les mêmes processus néologiques.

On trouve ainsi de nombreux mots *vha* qui n'ont pas d'équivalent en islandais, et inversement (par exemple *vha. boum* / v.isl. *tré* ou v.isl. *drengr* / *vha. kneht*) ou de mots néerlandais qui divergent de ceux de l'allemand (*achter* / *nach*, *sprache* / *taal*, etc.). En d'autres termes, un étymon francique repose sur un hiatus méthodologique : s'il a existé sous une forme correspondante à celle du vieux haut-allemand, alors il est probable que la forme reconstruite est correcte. Mais rien ne dit qu'il n'a pas existé sous une forme lexicalement différente, ni que le mot *vha* a eu un équivalent phonétique en vieux francique. Reconstruire un étymon francique à partir du *vha* ou du vieil islandais relève donc du pari, *a fortiori* quand le vieux et le moyen néerlandais ne sont pas consultés. Cette fragilité de l'étymon francique mérite d'être soulignée en regard de l'étymon gaulois, qui lui est généralement attesté, d'une part, et le précède historiquement d'autre part.

L'argument cardinal en faveur des étymologies franciques est un syllogisme : 1) si un mot français semble apparenté à un mot germanique au sens large, alors son étymologie est sans doute germanique ; 2) le francique est le vecteur historique des lexèmes germaniques en français ; 3) donc l'étymon recherché est francique. Mais si un étymon reconstruit en vieux-francique est en concurrence avec un étymon attesté en gaulois, il n'y a aucune raison de retenir l'étymon francique.

En théorie, la phonétique historique permet de retrouver l'origine de l'emprunt, à condition de connaître soit l'étymon proto-Indo-Européen original, soit une proto-forme commune au celtique et au germanique et à la condition qu'une évolution propre au germanique ou au celtique soit observable dans le lexème étudié (par exemple, la disparition du **p* en proto-celtique ou les effets des lois de Grimm et de Verner en proto-germanique). Ces conditions sont malheureusement rarement réunies, et la phonétique ne permet pas souvent de trancher – voir à ce sujet les correspondances germano-celtiques dans Lane(1933) et Quentel (2012).

Dans le but d'illustrer ces différents problèmes méthodologiques, nous allons à présent étudier deux cas de mots auxquels les dictionnaires assignent habituellement une origine francique.

chouette : Le mot est échoïque selon FEW qui précise « d'après le cri de l'oiseau ». Selon le TLF, il s'agirait d'un hybride lexical issu de la forme a.fr. *çute* et de la forme a.fr. *choe*, une forme picarde issue du francique **kawa* reconstruite à partir du moyen néerlandais *couwe*, du norvégien *kaie* et du suédois *kaja*. Seul l'italien comporte une racine analogue, *civetta*, mais il s'agit d'un emprunt au français, et il est en concurrence avec *gufo*. Les autres langues romanes proposent toutes des racines différentes ; portugais *coruja*, espagnol *lechuza*, occitan *machòta*, lazio *arpè*, roumain *bufnă*, sarde *cucumèa* etc.

La proposition du TLF est contestable car elle se base sur l'hybridation d'un échoïsme, qui ne peut pas être attesté d'une part et qui ne reproduit en aucun manière le cri de chouette d'autre part (laquelle ne chuinte pas mais (*h*)*ulule*, terme germanique basé sur la série *uil / owl / ul / eule* qui désigne justement l'oiseau en question). On notera incidemment que l'italien *gufo* est lui aussi supposément échoïque...

Du point de vue sémantique, l'étymon germanique supposé **kawa* pose un autre problème : il ne désigne pas une chouette mais un choucas (de même que les mots *couwe*, *kaie* et *kaja* sur lesquels il s'appuie). Ce sont des oiseaux tout à fait différents, appartenant à la famille des corvidés et qui ne présentent aucune similitude de forme, de couleur ou de comportement avec le rapace nocturne qui nous intéresse (Temminck, 1840 : 89). Or il existe un gaulois *caouanos* attesté dans une glose (cf. DELL, FEW et LEW) et dans la toponymie (Chavenay, de *Cavenoilus* : « la clairière de la chouette ») (DLG). Cette forme se retrouve dans le breton *kaouenn*, qui désigne également une chouette. Ici, l'étymon gaulois paraît nettement préférable à l'étymon francique, car il est 1) phonétiquement aussi acceptable, 2) attesté alors que l'autre ne l'est pas, 3) sémantiquement conforme à son aboutissement alors que l'autre ne l'est pas et que 4) le caractère germanique de la racine *kawa* n'est pas établi (elle est inexistante en vieux norrois, en gothique et en vha).

blet / blesser : Un autre cas intéressant de concurrence francique / gaulois : les mots « *blet* » et « *blesser* ». « *blet* » est habituellement rattaché à « *blesser* », d'après un radical frq. **blettjan*, « meurtrir » [vha *bleizza*] selon Rob. Selon FEW, « *blet* »

serait lié au vieil anglais *blât* « blême » et au vieux haut allemand *bleizza* « tache bleue produite par une meurtrissure ». Cette explication est spéculative car en vha *bleizza* signifie simplement « bleu » : le dictionnaire de Köbler ne donne aucune autre entrée et il n'y est jamais fait allusion à l'idée de meurtrissure.

Le suffixe *-ette* est ajouté sur le modèle des adjectifs en *-et* / *-ette* (TLF). On connaît parallèlement le gaulois *blati*, « mou », attesté dans l'onomastique (*Blattius*, *Blattio*, *Ando-blationi*) et apparenté au vieil irl. *mláith* ainsi qu'au bret. *blot* (de même sens) (Delamarre, 2003 : 78). On remarque que le sens premier de *bletier* / *blecier* en AF est « amollir en battant » (Greimas, 2001 : 67), sens que l'on retrouve en suisse roman avec « blyési », « amollir » (TLF). Il ne se trouve pas de cognat dans les autres langues romanes, ni dans les langues germaniques au sens de « blesser » (danois *såre*, allemand *verletzen*,) ou d' « amollir ». On trouve également le danois *blød* (all. *bloss*), issu d'un proto-germanique **blauta* qui signifie lui aussi « mou » et même parfois « trop mûr » (Gyldendals), qui est de toute évidence apparenté au gaulois *blati*, bien que les dictionnaires étymologiques n'en fassent pas mention et leur proposent même des étymons PIE différents.

Phonétiquement : le francique **blettjan* est régulier. Le gaulois *blat* l'est aussi dans la mesure où la géminée [tt] est attestée dans les NP donnés par Xavier Delamarre (cf. ci-dessus) et sous une forme affaiblie dans le cornique *bleudhder* (douceur, tendresse).

Il existe d'autres cas de concurrence de ce type (par ex. *saule*, serait issu du francique **salha*, mais il existe un gaulois *salico* attesté), qui témoignent assez du fait que l'étymon gaulois est souvent négligé à tort. Nous n'avons pas retenu de cas basés sur des mots celtiques reconstruits à dessein afin de nous limiter à des formes gauloises attestées.

2.3. LE CELTIQUE

Les langues celtiques se caractérisent phonétiquement par l'effacement de la labiale indo-européenne **p* (cf. lat. *pater* / gaul. *ater*). Ceci entraîne que tout mot celtique issu du PIE doit se conformer à cette loi, sans quoi il est considéré comme non-celtique. La famille celtique se subdivise à son tour en deux branches sensiblement éloignées l'une de l'autre, et dont la caractéristique la plus remarquable est le passage de la labio-vélaire indo-européenne **k^w* à la vélaire *k* dans la branche gaélique et à la labiale *p* dans la branche brittonique (cf. bret. *penn* et irl. *cenn* « tête ») (MacAulay, 1992 : 3). On notera que le celtique brittonique partage cette dernière particularité avec les langues sabelliques (Clackson & Horrocks, 2007 : 49). En termes étymologiques, cette loi permet d'affiner la précédente car à un mot brittonique en *p* doit correspondre un mot gaélique en *c* ([k]) pour disqualifier l'hypothèse d'un emprunt au

latin ou au germanique. Cette seconde loi fonctionne également pour des mots non-IE. Le gaulois est considéré comme très proche du groupe brittonique, dont il partage de nombreuses caractéristiques phonétiques et lexicales (Lambert, 2003 : 14 ; Abalain, 2000 : 22) ainsi que la mutation $*kw \rightarrow p$, entre autres.

La reconstruction d'étymons gaulois se fait donc principalement sur la base des langues brittoniques modernes et / ou par comparaison avec le latin, mais il faut admettre qu'elle se fait rarement en-dehors du strict domaine des études celtiques, dans le but de retrouver des étymologies dans les langues celtiques elles-mêmes. Ceci est un autre paradoxe : nous disposons de sources à la fois plus vastes et parfois bien plus anciennes pour les langues celtiques que pour les langues germaniques, le gaulois est un substrat du français, mais l'on trouve peu d'étymons gaulois reconstruits dans les dictionnaires, où l'on fait à l'inverse un usage intensif de l'étymon francique reconstruit. On peine à s'en expliquer les raisons.

Ardoise, en constitue une intéressante illustration : du celtique *ard*, « haut » selon FEW. La racine gauloise *ard-* serait également présente dans le mot « Ardennes » (du très ancien toponyme gaulois *ercunia* : forêt de chênes). Le TLF reprend l'étymologie du FEW et précise que le terme est spécifique au nord de la France. La reconstruction phonétique apportait ici une réponse plus satisfaisante puisqu'on trouve un mot gaulois reconstruit (Delamarre, 2003 : 55) **aritisia* apparenté au latin *paries*, *-etis* et que P-Y Lambert (2003 : 189) reconnaît comme l'étymon du français « ardoise », lequel est traditionnellement associé au NP *Ardennes* sans raison valable : les principaux gisements ardoisiers français ne sont pas dans les Ardennes et ne l'ont jamais été (Smeyers, 1958 : 90).

2.4. LES LANGUES SUBSTRATIQUES PRÉ-INDO-EUROPÉENNES

Outre le mot « ardoise » qui vient d'être mentionné, il existe un certain nombre de mots communs au gaulois et au français ou au français et au breton dont on ne retrouve aucun cognat dans les autres langues celtiques ou romanes. Si le breton est une langue brittonique, il n'en demeure pas moins qu'elle a absorbé une quantité non-négligeable de mots gaulois ou pré-celtiques locaux que l'on ne trouve pas en celtique insulaire (Abalain, 2000). Les langues celtiques en général possèdent un stock très important de lexèmes non-indo-européens dont l'origine est sans doute substratique. Il existe une série de règles permettant d'identifier les mots d'origine pré-IE (donc pré-celtique dans le cas du français), voir à ce sujet Salmons (2004). Il est évidemment impossible d'avoir la certitude définitive qu'un mot a une origine substratique en raison de l'absence de traces attestées de ces langues, mais l'on peut raisonnablement signaler les cas où cette hypothèse est probable. En voici deux exemples qui mettent en jeu l'hypothèse celtique, et la méthode proposée pour démêler l'écheveau de ces étymologies :

Roche : L'origine de ce mot est de l'avis général (TLF, Robert, OED...) non élucidée. On le trouve dans les langues du groupe gallo-roman (*rocca, roca, roccia*, etc.) ainsi qu'en breton (*roc'h*). L'anglais *rock* est un emprunt au français. En raison de l'existence du mot breton *roc'h*, OED suggère que le mot a une origine celtique, mais on ne le retrouve dans aucune autre langue du groupe celtique que le breton, ce qui indique que soit 1) le terme a disparu de ces langues soit 2) il est propre au celtique continental, et donc d'origine pré-celtique (substratique). Le fait qu'il ne se trouve aucun cognat de ce mot en roumain ni en sarde exclut une origine latine. Il n'existe aucun étymon PIE de ce mot.

Parc : Les notices des dictionnaires divergent nettement sur ce mot. On le retrouve dans les langues du groupe Gallo-Roman (it. *parco*, esp. *parque*), dans la racine du latin *parcere* (« retenir »), dans les langues celtiques (bret + corn. + gall : *park*, irl. *pairc*) et dans le Germanique de l'ouest (all. *Pferch*, vha. *pfarrih*, néerl. *park*). Pour OED, il est issu du proto-germanique **parruk*, mais le TLF conteste cette opinion en raison de la présence du [p] à l'initiale, quasi-inexistant en proto-germanique car issu d'un PIE **b* dont l'existence même semble improbable (voir aussi Diez, 1871 : 299) et du suffixe latin. De son côté OED rejette la possibilité d'une origine celtique, là encore en raison de la présence d'un [p] initial (comme on l'a vu plus haut, le **p* PIE disparaît dans les langues celtiques). Ajoutons que si le terme avait été celtique, le [p] brittonique alors issu d'une labio-vélaire **kʷ* aurait dû avoir un équivalent en [k] en gaélique (graphié « c »). L'identité des [p] en gaélique et en brittonique exclut donc l'origine celtique. Ni germanique, ni celtique : reste l'hypothèse romane, confortée par la présence d'un latin *parcere* (retenir), mais infirmée par l'absence de toute trace de ce terme en roumain et en sarde. Il ne reste donc plus que l'hypothèse substratique, suggérée du reste par le TLF qui indique que « le mot est à l'origine un pré-latin **parra*, „perche” dont les dérivés se retrouvent sur tout le territoire gallo-roman ». Les mots celtiques et germaniques de l'ouest seraient donc des emprunts à ce substrat pré-latin, ce qui expliquerait alors la présence d'un *p* initial en celtique et en germanique occidentale, ainsi que l'absence du mot dans les autres langues des familles germanique et celtique.

3. PROBLÈMES MÉTHODOLOGIQUES

Outre la question des interférences substratiques qui peuvent masquer ou au contraire suggérer à tort une origine gauloise, il existe un certain nombre de problèmes d'ordre strictement méthodologique qui leur sont directement liés. Ces problèmes ne concernent pas seulement le gaulois, car s'ils amènent à considérer un étymon comme latin ou francique sur la base d'arguments très fragiles. Ils peuvent dissimuler toute autre origine qui aurait méritée d'être considérée. Nous passerons succinctement ces problèmes en revue.

3.1. LES DISTORTIONS SÉMANTIQUES

La distortion sémantique consiste à altérer, parfois de façon considérable, le sens d'un mot pour le faire coïncider avec un étymon supposé. C'est un procédé auquel l'étymologie a souvent recours : il est parfois attesté, mais le plus souvent il ne l'est pas. Il se manifeste notamment dans des cas comme *bille* (supposément issu du francique **bikkil* : dé), *marcher* (du francique **markon* : marquer), *chômer* (du bas latin d'origine grecque *cauma* : forte chaleur), *couper* (de « coup » avec le sens de « diviser d'un coup »), *chenille* (du latin populaire **canicula* : petite chienne), et la liste des curiosités de ce genre est longue. Elle a d'ailleurs fourni l'essentiel des mots du *Dictionnaire des Etymologies Obscures* de Guiraud (1984). Pour illustrer l'occultation possible d'un étymon gaulois par ce procédé, l'on examinera le cas du mot « lisse » qui n'offre de correspondance ni sémantique ni phonétique avec les étymons qu'on lui assigne :

lisse : Selon le TLF et FEW : croisement du latin *lixare* : faire cuire dans l'eau avec *allisus* : élimé. L'analogie est risquée. Le mot se retrouve dans l'espagnol *liso*, l'italien *liscio*, le portugais *liso*. Or nous avons également le gaulois *lica* : pierre plate, attesté (Delamarre 2003), que l'on retrouve dans le breton *lec'h*, le gallois *llech* et le vieil irlandais *lecc*, mais aussi et avec le même sens dans l'occitan *lhéco*. La concordance phonétique est certes critiquable, car il aurait fallu avoir un étymon de type **licia* pour aboutir à « lisse », mais la concordance de sens et de forme, l'existence de formes *lici-* attestées dans la toponymie ainsi que l'absence de cognats en roumain sont des indices suffisants pour que l'on examine l'hypothèse sérieusement.

3.2. LES ÉTYMOLOGIES ÉCHOÏQUES

L'étymologie échoïque est un concept éminemment problématique dans le domaine de l'étymologie dans la mesure où il est impossible de l'attester et où, dans bien des cas (bobine, etc.) il semble relever du plus pur impressionnisme. Ajoutons que les onomatopées ne respectent aucune des lois de la phonétique historique (supposément pour conserver leur potentiel expressif), qu'elles contredisent l'idée fondatrice d'arbitraire du signe, que dans bien des cas (cris d'animaux notamment) on retrouve sans mal leurs étymons latins ou grecs (*aboyer*, *balbutier*, *beugler*, *coasser*, etc. voir à ce sujet Quentel 2011) et enfin qu'elles n'apportent aucune réponse à la question étymologique fondamentale qui est de savoir de quelle langue provient un mot donné. L'onomatopéique n'est pas une langue, c'est un processus néologique, et si processus il y a eu, il reste à savoir dans quelle langue il s'est produit. Ce concept parasite considérablement les études étymologiques, et ceci d'autant plus que les dictionnaires en abusent depuis les origines. Or il est à peu près clair que ces étymologies dissimulent les étymons réels qui, eux, attendent toujours d'être découverts.

Il en existe plusieurs que l'on a de bonnes raisons de soupçonner d'être gaulois. C'est notamment le cas des mots « petit », « coq », et « coquelicot » (cf. Quentel 2011).

3.3. LES ÉTYMOLOGIES IRRÉGULIÈRES

La régularité phonétique a longtemps été érigée comme l'élément déterminant de la reconstruction des étymons. La méthode phonétique est en soi un outil d'investigation très puissant dans le cadre des études étymologiques (Malkiel, 1993 : 22). Elle est pourtant de plus en plus contestée, non pas pour sa fiabilité, mais pour l'usage qui en a été fait : filiations démontrées phonétiquement aux dépens de la vraisemblance sémantique (*chômer*, *chenille*, *rue*, etc.), entorses régulièrement faites aux lois de la phonétique historique au nom d'exceptions dialectales (*cage*, *bœuf*, etc.) ou systémiques (*donjon*), si nombreuses qu'elles font vaciller tout l'édifice, créations d'étymons non-attestés (issus notamment du latin populaire et du francique) que l'on peut façonner à sa convenance, souvent au prix de quelques entorses, là encore, à la régularité phonétique (on l'a vu par exemple pour *ardoise* et pour *chouette*, mais les cas sont nombreux) et à la vraisemblance sémantique.

Ce problème est considérable dans la mesure où il permet de façonner des étymons quasiment à volonté, et cela d'autant plus facilement que la filiation sémantique n'est pas considérée comme une contrainte.

3.4. LE LATIN POPULAIRE

La question du latin populaire (LP) est éminemment complexe, et nous n'entrons pas ici dans les détails du sujet. Le problème que pose le latin populaire dans le cadre de la recherche étymologique est identique à celui que pose le francique : la langue n'est pas attestée. En outre, sa nature-même est mal définie, car toujours sujette à débats et à études parfois contradictoires : le latin populaire est-il fille ou sœur du latin classique (Mańczak, 1994) ? Dans quelle mesure a-t-il subi des influences étrangères ? Quel rôle ont joué les substrats (italiques, celtiques, ibériques, étrusque et en général pré-IE) dans sa formation ? Quelle était la fréquence réelle de son usage ? Comment a-t-il évolué et à quelle époque a-t-il été parlé ? Par qui (locuteurs natifs du latin ou pas) ? Sur quelle étendue géographique ? Avec quelles variations dialectales ? À quelle époque précise ? Toutes ces questions n'ont, pour l'heure, pas obtenu de réponse claire. À ceci s'ajoute le problème de la nature transitoire de cette langue : un mot qui n'a pas de cognat en latin classique (voir ci-dessous le cas de *faucon*) ne peut pas être considéré comme issu du latin, mais plutôt comme un emprunt d'une autre langue. La réponse à la question étymologique se trouve donc dans la langue prêteuse, et non dans le LP en lui-même. Malgré cela, les pages des dictionnaires sont pleines d'étymons LP non-attestés, créés pour la plupart par Gustav Gröber à la fin du XIX^{ème}

siècle (Malkiel, 1993 : 24 ; Gröber, 1889). Cela accroît d'autant les possibilités de façonner des étymons hors de toute contrainte, et le risque d'aboutir à des étymons erronés est très grand (le cas de **canicula* « petite chienne » → *chenille* relevé par Guiraud est emblématique de ce problème). Et bien entendu, un étymon LP erroné occulte l'étymon réel. Un mot comme *faucon* illustre bien ce problème.

Faucon : issu du bas latin *falco*, *-onis*, par analogie avec la forme du bec (FEW) ou des ailes (Robert). Seules les langues du domaine gallo-roman possèdent une racine de type *falco* (portugais *falcão*, espagnol *halcón*, occitan : *falcon*, italien *falco*, mais lazzio *arpè* (du latin *accipiter*). Il n'y a pas de cognat en sarde, en roumain ni en sicilien. Quant au lien sémantique, il se fonde sur une synecdoque non-attestée, et il n'est pas convaincant : il n'y a pas de lien objectif entre le bec ou les ailes d'un faucon et la forme d'une lame de faux. Le bec du rapace est « court et courbé depuis sa base » (Temminck, 1840 : 16), ses ailes déployées en vol ne présentent aucun courbure qui rappelleraient la forme de la faux. Pour le reste, le bas latin pose le même problème que le LP, bien qu'étant attesté : il ne donne pas la source étymologique du mot. Tout ce que l'on peut constater, c'est que le mot n'est pas latin (*accipiter*). Il peut donc s'agir soit d'un emprunt, soit d'un héritage d'une langue de substrat. C'est vers cette seconde hypothèse que l'on se dirigera, car il existe un gaulois *volcos* (faucon), attesté par l'onomastique (Delamarre, 2003 : 326) : *Volques Tectosages*, *Volques Arécomices*, *Uolcius*, *Uolcenius*, *Uolcaci*, *Catu-uolcus*. Le sens est identique, la phonétique régulière. On retrouve cette racine dans toutes les langues celtiques : breton *falc'hun*, cornique *falghun*, irlandais *fabhcún* et gallois *gwalch* : l'origine celtique du terme est ici confirmée par l'aboutissement à « c » en vieil irlandais et à « c'h »/« gh » en brittonique de la vélaire PIE **k*, ainsi que par l'aboutissement à « f » en vieil irlandais et à « gw » en gallois de la semi-voyelle PIE **w*, laquelle correspond régulièrement au « v » ([w]) du gaulois.

À titre indicatif, signalons les cas des mots *bassin*, supposément du LP **baccinus* (TLF) alors qu'il y a un gaulois **bascauda* qui laisse FEW perplexe mais que Lambert (2003 : 189) confirme comme étymon de « bassin », et celui de *carrière* du LP **quadraria* qui serait issu par contamination de *quadrus lapis* (une figure de composition sémantique qui mériterait que l'on s'y attarde : « forêt » de *silva forestis*, « sanglier » de *singularis porcus*, etc). Dans ce dernier cas nous n'avons pas d'étymon gaulois, mais une divergence de mots romans (*cava*, *cantela*, *pedreira*) et une convergence de mots celtiques : gallois *carreg*, breton *karreg*, cornique *karrek* et irlandais *creig*, lesquels désignent tous une pierre.

3.5. LES RÉEMPRUNTS

Il s'agit de mots gaulois ayant transité par le latin pour revenir vers le LP. Ces emprunts sont attestés par les auteurs antiques (Ausone, Pline, Polybe, etc.). C'est notamment le cas de *carros* (char), de *benna* (benne) ou encore de *lanca* (lance), *gladius*

(glaive), etc. Lambert (2003 : 204 et suivantes) mentionne de nombreux mots latins empruntés au gaulois. La phonétique indo-européenne permet en outre de détecter des cas d'emprunts plus anciens, tel que *bucca* (bouche), du gaulois *bocca* (Delamarre, 2003 : 78). Ce qui pose problème ici n'est pas tant le processus d'emprunt, mais plutôt le processus de réemprunt par les gallo-romains de mots gaulois qu'ils connaissaient déjà. Ce détour paraît défier la logique, car on se demande alors pourquoi la population de la Gaule qui a continué à pratiquer la langue gauloise pendant la période impériale (parmi d'autres, signalons que Sidoine Apollinaire constate au V^{ème} siècle que la noblesse arverne vient tout juste d'apprendre le latin), éprouverait le besoin d'emprunter un mot gaulois au latin... La phonétique ne permet guère de trancher : dans des cas comme *carros* / *carrus* elle n'est pas éclairante.

À l'inverse, quand l'étymon est un mot gaulois emprunté au latin, la transition n'est pas mentionnée. Delamarre signale ainsi que « le français *chétif*, provençal *caitiu* viennent eux aussi du latin *captivus* « prisonnier », mais par un intermédiaire gaulois **caχtiuos* avec *-pt-* → *-χt-* ; vieil emprunt des Celtes au latin ».(Delamarre, 2003 : 112).

4. CONCLUSION

L'étymologie gauloise en français pose un certain nombre de problèmes spécifiques au rôle des substrats linguistiques anciens dans la genèse des langues modernes, lesquels sont de deux natures : celle des rapports entre les strates historiques de lexicalisation qui s'occultent ou se parasitent les unes les autres, et celle de la méthodologie qui privilégie de façon souvent arbitraire une strate plutôt qu'une autre ou imagine des concepts tels que l'échoïsme ou le latin populaire qui non seulement n'apportent aucune réponse à la question étymologique mais en outre paralysent la recherche d'étymons réels.

Les rapports interstratifiés jouent toujours en défaveur du gaulois, pour des raisons que l'on peine à s'expliquer car, bien que peu attesté, il l'est quand même, alors que ni le francique ni le latin populaire ne le sont. De ce fait, des étymologies franciques ou latines populaires occultent des étymons gaulois pourtant évidents (*chouette*, *petit*, etc.). Dans le même ordre d'idées, plusieurs étymons gaulois sont en concurrence avec des étymons latins sans que cela soit jamais mentionné.

Il existe en outre un certain nombre de problèmes d'ordre strictement méthodologique qui sont liés aux précédents : altérations sémantiques (*chômer*, *chenille*, *rue*, etc.), phonétiques (*cage*, *bœuf*, etc.) ou systémiques (emploi de composition lexicale comme pour *donjon*) destinées à forcer la preuve d'un étymon germanique ou latin populaire, lequel n'est jamais attesté. A ceci s'ajoute encore le cycle prêt/réemprunt du celtique vers puis du latin qui paraît difficile à concevoir, et enfin le problème de la nature même du latin populaire, qui ne résout aucune question étymologique puisque

soit le mot LP est soit *in fine* latin (classique), soit il est issu d'une autre langue et dans ce cas il est nécessaire de découvrir laquelle : « latin populaire » n'est pas davantage une origine que « échoïque ».

Ajoutons enfin que, si les dictionnaires utilisent de façon extensive les étymons reconstruits en latin populaire et en francique, ils ne font guère usage d'étymons gaulois reconstruits. Cet article s'est limité aux étymons gaulois attestés : il y aurait pourtant matière à prolonger la recherche en s'intéressant aux étymons celtiques reconstruits.

Tout ceci ne remet évidemment pas en question le fait que l'essentiel du lexique français est latin (classique) : l'impact d'une langue de substrat est, en termes lexicaux, toujours mineur. Mais sa réévaluation systématique, pour le français comme, du reste, pour l'anglais, reste encore à faire.

RÉFÉRENCES

- ABALAIN, Hervé (2000): *Histoire de la langue bretonne*. Paris : Gisserot.
- ALINEI, Mario (2001): « The Celtic Origins of Latin Rota ». *Studi Celtici* III : 13-29.
- BAMMESBERGER, Alfred (1999): « Die Runeninschrift von Bergakker: Versuch einer Deutung ». In: Alfred BAMMESBERGER, Pforzen und Bergakker. Neue Untersuchungen zu Runeninschriften. *Historische Sprachforschung* 41: 180-185.
- BLASCO FERRER, Eduardo (2010): *Paleosardo. Le radici linguistiche della Sardegna neolitica*. De Gruyter : Berlin.
- BONNARD, Henri (1992): *Synopsis de phonétique historique*. Paris : Sedes.
- BOURCIER, Edouard (1930): *Précis Historique de phonétique française*. Paris : Klincksieck.
- CLACKSON James; Geoffrey HORROCKS (2007): *The Blackwell History of the Latin Language*. Malden : Blackwell Publishing.
- DE LA CHAUSSÉE, François (1974): *L'initiation à la phonétique historique de l'ancien français*. Paris : Klincksieck.
- DELAMARRE, Xavier (2003): *Dictionnaire de la langue gauloise*. Paris : Errance.
- DIEZ, Friedrich (1874): *Grammaire des langues romanes*, vol. II. Paris : Morel-Fatio.
- DOTTIN, George (1918): *La Langue Gauloise : grammaire, textes et glossaire*. Paris : Klincksieck.
- ESKA, Joseph F., David E. EVANS (1993): « Continental Celtic ». In : Martin J. BALL (éd.). *The Celtic Languages*. London : Routledge, 26-63.
- GREIMAS, Algirdas J. (2001): *Dictionnaire de l'ancien français*. Paris : Larousse.
- GRÖBER, Gustav (1989): *Vulgärlateinische Substrate romanische Wörter. Archiv für lateinische Lexikographie und Grammatik*. Strasbourg : S.N.L.
- GUIRAUD, Pierre (1983): *Dictionnaire des étymologies obscures*. Paris : Payot.
- KÖBLER, Gerhard (2006): *Neuenglisch-althochdeutsches Wörterbuch*. <<http://www.koeblergerhard.de/germanistischewoerterbuecher>> (consulté le 11/12/2011).
- LAMBERT, Pierre-Yves (2003): *La langue gauloise*. Paris : Errance.
- LANE, Geo S. (1933): « The Germano-Celtic Vocabulary ». *Language* 9-3 : 244-264.
- LEJEUNE, Michel (1971): « Documents gaulois et para-gaulois de Cisalpine ». *Études Celtiques* 12 : 357-500.
- MacAULAY, Donald (1992): *The Celtic Languages*. Cambridge : Cambridge University Press.
- MEES, Bernard (2002): « The Bergakker Inscription and the Beginnings of Dutch ». In : Theo VEN-NEMANN, *Amsterdamer Beiträge zur Alteren Germanistik* 56. Amsterdam : Rodopi, 23-26.

- MALKIEL, Yakov (1993): *Etymology*. Cambridge : Cambridge University Press.
- MIHAILA, Gheorghe (1980): « Les attestations les plus anciennes des mots autochtones en roumain ». In : *Actes du I^e Congrès international de thraçologie*. Bucarest : Editura Academiei, 75-84.
- MAŃCZAK, Witold (1994): *Proto-Roman et Origines des langues romanes*. Amsterdam : Benjamins.
- POKORNY, Julius (1966): *Indogermanisches etymologisches Wörterbuch*. Bern : Francke.
- QUENTEL, Gilles (2011): « La Néologie Onomatopéique dans l'étymologie du français ». *Revue Romane* 46 (1) : 89-104.
- QUENTEL, Gilles (2012): « Early Linguistic Contacts between Celtic and Germanic : Lexical Aspects ». In : Andrej KAŤNY (éds.), *Sprachkontakte in Zentraleuropa*. Frankfurt am Main : Peter Lang Edition, 175-185.
- SALMONS, David (2004): « How (non-)Indo-European is the Germanic Lexicon? And what does that mean? ». In : Irma HYVÄRINEN, Petri KALLIO, Jarmo KORHONEN (éds.) *Etymologie, Entlehnungen und Entwicklungen: Festschrift für Jorma Koivulehto zum 70. Geburtstag*. Helsinki : Société Néophilologique, 311-332.
- SMYERS, Louis (1858): *Essai sur l'état actuel de l'industrie ardoisière en France*. Paris : Poulet-Malassis & De Broise.
- TEMMINCK, Coenraad J. (1840): *Manuel d'ornithologie*. Paris : Dufour.
- UNTERMANN, Jürgen. (1997): *Monumenta Linguarum Hispanicarum. IV Die tartessischen, keltiberischen und lusitanischen Inschriften*. Wiesbaden : Reichert.
- ZINK, Gaston (1986): *Phonétique historique du français*. Paris : PUF

SOURCES ÉLECTRONIQUES

Source électronique 1 :

- KÖBLER, Gerhard (2006): *Neuenglisch-althochdeutsches Wörterbuch*. <<http://www.koeblergerhard.de/germanistischewoerterbuecher>> .Consulté le 11 décembre 2011.