

Le programme TLF-Étym: apports récents de l'étymologie comparée-reconstruction

Gilles Petrequin
Marta Monda Andronache
Centre National de la Recherche Scientifique et Université de Nancy

Our topic is the French hereditary vocabulary considered in a new theoretical approach and focuses on the theme "Historical and learned Lexicography and Etymology". From the point of view of the classical etymology any lexeme must find its origin in a graphic form. Therefore, there is nowadays a strong consensus among the specialists of the Romanic studies to revise this "classical" and philological lexicographic practice that puts the graphic form at the centre of the theory, which generates basic contradictions. Recognizing the oral form of a hereditary lexeme and renouncing the "graphic centrist" conception in the treatment of the hereditary vocabulary appears as an obvious necessity in the daily practice of the lexicographer. Moving away the "classical" method of the Romanic etymology, we propose to apply the system of the historical and comparative grammar to the field of the French etymology to reconstruct, by comparing different oral forms from Romanic languages, the oral form of the proto-language. Our submission presents three examples of etymological notes/headwords on the hereditary vocabulary developed and published by the program TLF-Étym of the linguistic laboratory ATILF (Analyse et traitement informatisé de la langue française; CNRS/Nancy-Université, France). These examples will allow us to demonstrate to what extent the practice of the etymology of the French hereditary word pool depends on the progress of the Roman etymology, with which it should go hand in hand from now on.

1. Présentation du programme TLF-Étym

Le programme de recherche TLF-Étym constitue l'une des actions de recherche de l'équipe *Linguistique historique française et romane*, elle-même composante du laboratoire *Analyse et Traitement Informatique de la Langue Française* (ATILF), situé à Nancy (France)¹. Actuellement placé sous la direction de Gilles Petrequin, les membres du programme TLF-Étym se sont donné pour tâche, avec le concours d'une trentaine de lexicologues francisants du monde entier, de réviser sélectivement les notices intitulées "Étymologie et Histoire" du *Trésor de la langue française informatisé* (TLF(i), cf. <http://www.atilf.fr/tlfi.htm>). Il ne s'agit pas de réécrire l'ensemble des 54.280 notices historico-étymologiques du TLF(i), mais de revoir plus systématiquement un ensemble de notices à partir des trouvailles faites par les différents chercheurs dans leur activité propre, d'exploiter une source nouvellement publiée et, plus globalement et plus fondamentalement, de remédier à des erreurs de méthode de l'ouvrage à ses débuts (cf. Buchi 2005).

Le lancement du projet —par Éva Buchi en janvier 2005—, a été motivé par trois souhaits. Premièrement, enrichir et corriger la partie étymologique du TLF(i), qui conserve une fonction de divulgation, auprès du grand public cultivé, des acquis de la lexicologie/lexicographie française. Deuxièmement, s'adapter au caractère éminemment progressif de la recherche scientifique et répercuter les résultats de la linguistique contemporaine. Et, troisièmement, mettre à profit les nouvelles possibilités qui s'offrent à la lexicographie du XXI^e siècle grâce à l'informatique: le passage du TLF ("version papier") au TLFi ("version informatisée") a constitué un immense pas en avant du fait des multiples options d'interrogation transversale et multicritère

¹ ATILF (CNRS/Nancy-Université, 44, avenue de la Libération, BP 30687, F 54063 Nancy cedex, France. <http://www.atilf.fr/> [accessible en mars 2008].

qu'offre le moteur de recherche STELLa par la diversité de ses champs d'interrogation². Il s'agit donc pour le programme TLF-Étym, à la fois d'ancrer solidement le TLF(i) dans la lexicographie évolutive (Martin 1998) et de profiter de l'avantage manifeste que présente l'outil informatique sur l'outil papier: la facilité relative des corrections et des mises à jour de l'information.

Le programme TLF-Étym s'appuie non seulement sur les résultats des recherches menées par les membres qui constituent son équipe de rédacteurs mais aussi sur l'ensemble des méthodes de travail que mettent en œuvre les autres composantes de l'équipe *Linguistique historique française et romane*, qu'il s'agisse notamment de la méthode historico-comparative (FEW, dirigé par Jean-Paul Chauveau)³ ou de l'analyse philologique des textes anciens (DMF, dirigé par Sylvie Bazin-Tacchella)⁴.

D'ores et déjà (mars 2008) la base TLF-Étym, consultable en ligne (<http://www.atilf.fr/tlf-etym>), contient près de 150 notices étymologiques révisées et plus de 1120 notices bibliographiques. Les améliorations apportées aux anciennes notices du TLF sont de diverses natures: étymologies inédites, précisions étymologiques, antédations de lexèmes, rétrodatations de lexèmes, antédations de sens secondaire. L'apport essentiel est cependant d'ordre méthodologique, ce sur quoi nous allons insister dans cet article, en prenant pour exemple le cas du lexique héréditaire.

2. Cadre méthodologie général du programme de recherche TLF-Étym

Le cadre méthodologique de TLF-Étym est constitué par ce que Kurt Baldinger (1959) appelle "l'étymologie-histoire du mot" et Pierre Guiraud "l'étymologie lexico-historique": une conception qui part du principe que l'étymologie ne se ramène pas au simple énoncé d'un étymon, mais qu'elle a vocation à établir explicitement le rapport formel et sémantique entre ce dernier et le lexème à expliquer, et que l'histoire des mots fait partie intégrante de l'étymologie. Les origines de cette conception remontent à Schuchardt, Meillet, Ernout et surtout à von Wartburg; mais c'est Kurt Baldinger (1959) qui l'a sans doute le mieux formulée: "l'étymologie, [...] c'est [...] la biographie du mot".

Au sein de ce cadre théorique général, chaque lexème, eu égard à la classe étymologique dont il relève (vocabulaire héréditaire, emprunts, créations françaises), doit faire l'objet d'une démarche analytique particulière qui met en œuvre une méthodologie propre à la classe étymologique concernée. La catégorie des lexèmes héréditaires pose des problèmes spécifiques d'analyse, et c'est celle dont la méthodologie a été récemment le plus considérablement renouvelée.

3. L'étymologisation des lexèmes héréditaires: un problème méthodologique particulier

La méthodologie appliquée dans les notices "Étymologie-Histoire" du TLF(i) consacrées aux lexèmes héréditaires est entièrement conforme à celle qui fut développée dans les travaux de tous les grands romanistes du XX^e siècle (cf. Chambon 2007: 58-59) et telle qu'elle fut appliquée dans les manuels (cf. Bourciez 1967⁵ [1910¹]) et les grands dictionnaires classiques traitant de l'étymologie romane, le REW et le FEW (cf. Büchi 1996) notamment. Selon cette conception traditionnelle, tout lexème français héréditaire est supposé, idéalement, se rattacher à un étymon latin, attesté soit à l'époque classique soit à l'époque tardive. La base épistémologique de cette démarche rétrospective est foncièrement de nature philologique: les textes latins, qu'ils soient classiques ou tardifs, sont censés fournir les occurrences qui permettent d'asseoir l'étymon à partir duquel dérive le lexème français. Dans le cas où aucune attestation textuelle ne peut être avancée, l'habitude a été prise de reconstituer une forme latine écrite, avec au besoin les quantités vocaliques "jugées pertinentes pour rendre compte des

² Cf. les onglets "recherche assistée" et "recherche complexe" de la page d'accueil du TLFi (<http://atilf.atilf.fr/tlf.htm>).

³ Cf. http://www.atilf.fr/scripts/mep.exe?HTML=mep_few.txt [accessible en mars 2008].

⁴ Cf. http://www.atilf.fr/scripts/mep.exe?HTML=mep_mf.txt [accessible en mars 2008].

formes (gallo)romanes” (Chambon, à paraître). Qui plus est, lorsque aucune attestation satisfaisante n’a pu être trouvée en latin dit “classique”, on reconstruit alors une forme appartenant au latin dit “vulgaire”, que l’on affecte d’un astérisque pour signifier que cette forme n’est pas (ou pas encore) attestée dans une source écrite.

Le corollaire de cette démarche essentiellement philologique et qui s’intéresse avant tout à l’étymologie-origine, démarche adoptée nous l’avons dit par les rédacteurs du TLF, est que les lexèmes héréditaires français sont décrits, dans de nombreux articles de ce dictionnaire, comme “empruntés au latin” de l’Antiquité⁵, celui-ci étant appréhendé, selon la terminologie du TLF, à la fois selon les axes diachronique (latin [classique]/bas latin ou latin tardif), diaphasique ([latin classique]/latin vulgaire, latin populaire) et parfois sociolinguistique (langue vernaculaire/langue véhiculaire: latin ecclésiastique ou latin chrétien). C’est ainsi, par exemple, que, pour TLF(i), le français *agneau* (*anel* XI^e s., *a(i)gnel* au XII^e s.) “petit d’une brebis”, mot héréditaire qui possède des équivalents dans la plupart des langues romanes (REW 284), est simplement décrit comme un emprunt au lat. *agnellus* “petit agneau”, diminutif du lat. *agnus* “agneau”. C’est ainsi encore que TLF(i) donne pour étymon au français *asseoir*, autre lexème héréditaire dont les équivalents sont bien attestés dans d’autres parlers de la Romania (REW 7780), un supposé “latin vulgaire” **adsedere*, non attesté, lui-même “réfection du latin classique” *adsidēre* d’après *sedēre*; *sedere* “être assis”.

Cette méthode traditionnelle d’étymologisation du lexique héréditaire français, pour qui l’essentiel est d’indiquer l’origine lointaine du mot sans considérer la voie suivie et les processus de l’histoire linguistique, a pour inconvénient de mettre au centre de sa problématique la nature graphique du signe linguistique. Par ce fait même, on est alors conduit à nier la nature propre d’un lexème héréditaire, savoir qu’il est avant tout, et “nécessairement, un mot oral” (cf. Chambon, à paraître), qui s’est transmis verbalement pendant des siècles jusqu’au moment où il apparaît, fortuitement, dans des sources écrites, sous une ou plusieurs graphies qui ne rendent d’ailleurs pas fidèlement sa réalisation orale. Deux articles programmatiques récents dans le domaine de la linguistique romane de Jean-Pierre Chambon (2007 et à paraître) amènent à changer radicalement l’étymologisation des lexèmes héréditaires français (et plus globalement gallo-romans) et à adopter une démarche qui explicite la procédure souvent suivie intuitivement par les linguistes diachroniciens et qui rend compte d’une façon plus rigoureuse de l’histoire et de la constitution du lexique héréditaire français.

4. Nouvelle méthodologie dans l’étymologisation du lexique héréditaire

Très légitimement, la méthodologie de TLF-Étym s’est démarquée, dans certains domaines, des options qui avaient été celles des rédacteurs du TLF, qu’il s’agisse de la terminologie ou des procédures d’analyse mises en œuvre. Concernant le lexique héréditaire, il était donc également tout naturel que TLF-Étym s’inscrivît dans le mouvement de renouvellement des études romanistes qui s’observe à l’heure actuelle⁶ et qu’il en rende compte à sa manière, en

⁵ En toute rigueur, il convient de dire que nous adoptons le point de vue d’un lecteur actuel du TLF —peut-être un peu naïf —, qui prend les informations transmises au pied de la lettre et qui donne aux termes métalinguistiques (ex. “emprunté [au latin]”) leur sens le plus obvie. Il est vraisemblable qu’une étude historique et critique de la métalangue du TLF démontrerait que le défaut de l’analyse linguistique dans de nombreux articles est due à la fois au manque de réflexion théorique et à l’imprécision de la terminologie employée. Les rédacteurs du TLF désignaient, semble-t-il, l’*etimologia remota* d’un lexème héréditaire par le terme *emprunt*, en négligeant les étapes intermédiaires ou la voie réellement suivie par le lexème. De ce fait, la distinction entre héréditaire et emprunté est neutralisée dans le discours des notices du TLF par manque d’intérêt pour les processus de l’histoire linguistique.

⁶ Voir le projet DÉRom lancé officiellement en janvier 2008 par Éva Buchi et Wolfgang Schweickard. Les grandes lignes du projet international DÉRom (*Dictionnaire Étymologique Roman*) ont été présentées au *XXV^e Congrès International de Linguistique et de Philologie Romanes* d’Innsbruck (cf. Buchi; Schweickard 2008). Dans une première phase, DÉRom a pour ambition de reconsidérer les quelque 500 bases étymologiques communes à l’ensemble des langues romanes (cf. Fischer 1969) et d’en présenter l’analyse phonologique, sémantique et historique. Cf. <http://www.atilf.fr/derom> [accessible en mars 2008].

modifiant au besoin sa propre pratique, au fur et à mesure de l'avancée des connaissances aussi bien théoriques que pratiques.

La reconnaissance et la prise en considération de la nature foncièrement orale d'un lexème héréditaire fait apparaître toute la contradiction que constitue le fait de donner à un mot oral un étymon latin dont les caractéristiques s'inscrivent avant tout dans la tradition graphique. Il s'agira donc désormais dans les nouvelles notices de TLF-Étym —et sur le modèle de ce qui se pratiquera dans le DÉRom (cf. Buchi; Schweickard 2008)— de renoncer à cette conception graphocentrique de l'étymologie des lexèmes héréditaires en adoptant le point de vue des locuteurs romans et donc, en dernière analyse, en tenant compte du fonctionnement du système linguistique de la langue dont on cherche à étymologiser les unités lexicales.

Ce changement de perspective induit comme conséquence que la recherche de l'étymologie d'un lexème héréditaire ne se justifie plus en partant, en amont, d'un supposé ancêtre latin, c'est-à-dire en cherchant le *devenir* d'un mot latin écrit (que celui-ci appartienne à l'époque classique ou à la latinité tardive, à la langue littéraire ou au latin dit vulgaire), mais en reconstruisant de manière régressive et par comparaison intra-romane, l'étymon oral qui a constitué la forme ancestrale des unités lexicales attestées dans les diverses langues romanes (cf. Hall 1950; Fischer 1969). Cette nouvelle démarche heuristique oblige l'étymologiste à faire coïncider, le plus rigoureusement possible, l'ensemble des informations que fournissent les diverses langues romanes, et à ne se contenter pas de rechercher les étymons du vocabulaire héréditaire d'une seule langue en raisonnant par la simple phonétique historique. Adoptant donc les méthodes de la grammaire comparée-reconstruction (cf. Chambon 2007: 67-69), l'étymologie des lexèmes héréditaires fait pleinement et explicitement sienne la méthode historico-comparative qui permet une reconstruction interne de protoformes romanes orales qui forment autant d'étymons communs à la plupart des langues romanes attestées à l'époque historique. Cette démarche théorique induit des conséquences pratiques dans la notation des phénomènes décrits. La forme ancestrale d'un lexème héréditaire, "mot oral" étant nécessairement elle-même de nature orale, elle devra être donnée sous sa forme orale reconstruite, présentée en notation phonétique, précédée d'un astérisque qui signale que l'étymon est reconstruit. Cet *astérisque de reconstruction* d'un "mot oral" n'est donc pas à confondre avec celui utilisé traditionnellement par les philologues et qui dénote le caractère non textuellement attesté d'un lexème (cf. Chambon, à paraître).

Il ne s'agit pas, naturellement, d'une pure reconstruction phonique, puisque la reconstruction repose, dans la plupart des cas, sur une interprétation des graphies, mais il s'agit précisément de dépasser le figement quelque peu artificiel des formes textuelles. Il ne s'agit pas non plus d'opposer reconstruction et philologie, mais d'instaurer une sorte de mouvement dialectique entre les deux procédures: l'étymologiste reconstruit grâce à une interprétation, à toutes les étapes de son travail, de la forme phonique recouverte par les graphies, puis s'assure —quand cela est possible— que sa reconstruction est compatible avec la documentation établie par la philologie.

5. L'application pratique dans TLF-Étym de la méthodologie nouvelle

Nous présenterons ici trois exemples d'étymologie-reconstruction, *claie*, *alose* et *bigler*, chacun de ces lexèmes illustrant un cas différent, et se distinguant aussi par le degré de certitude auquel on peut atteindre par la méthode présentée ici.

Claie

Dans la version du TLF (5: 867), la notice "Étymologie-Histoire" du lexème *claie* donnait les informations suivantes⁷:

⁷ L'explicitation de l'ensemble des références bibliographiques citées dans les notices révisées de TLF-Étym sont disponibles sur le site de TLF-Étym, sous la rubrique bibliographie. Les notices bibliographiques sont établies à l'ATILF par les soins de Pascale Baudinot.

Étymol. et Hist. XI^e s. *cleide* judéo-fr. (*Gloses de Raschi*, éd. Darmesteter et Blondheim, t. 1, p. 27); 1155 *cleie* (WACE, *Brut*, 13546 ds KELLER, p. 283b); 1306 *claie* (G. GUIART, *Royaux Lignages*, I, 4012, *ibid.*); 1690 (FUR.: Claye est aussi une grosse échelle de charpente [...] sur laquelle on fait traîner par la ville ceux qui ont été tués en duel). Du gaul. *clēta* (DOTTIN, p. 246), attesté en lat. médiév. avec la graphie *clida* au sens de “treillage de bois sur lequel on déposait le cadavre d’un supplicié ou de la victime d’un meurtre” (VII^e s. *Lois Ripuaires* ds NIERM.).

Ainsi, TLF, à la suite de FEW (2: 778a, s.v. *CLETA) décrivait sommairement le lexème français *claie* comme un terme gaulois —*clēta*, forme non attestée philologiquement mais reconstruite à partir des formes conservées en irlandais, en gallois et en français (Dottin 1920: 246)— transmis par un texte latin d’époque mérovingienne (*Lois Ripuaires*, 7^e siècle), le lexème français constituant le représentant héréditaire de ce mot latin médiéval. Cette étymologisation de *claie* se justifiait aux yeux du rédacteur de la notice par le fait que la forme latine (*clida*) est attestée antérieurement aux formes romanes écrites (*Gloses de Raschi* 11^e s.), de sorte que, paradoxalement, comme le note Jean-Pierre Chambon (à paraître sous 6.1) à propos d’un cas similaire, “la donnée seconde [le latin] devient première et se trouve dotée d’un illusoire pouvoir explicatif. Les rapports historiques réels sont donc bien purement et simplement inversés par la perspective graphocentrique; le latin écrit emprunteur devient la source des formes héréditaires romanes”.

À la suite des travaux de Jean-Paul Chauveau et de Jean-Pierre Chambon, la perspective étymologique a été renversée dans la nouvelle version de TLF-Étym. En s’appuyant sur les données romanes disponibles (cf. déjà REW 1935³, 1988), il a été procédé à la reconstruction d’une forme ancestrale orale protoromane, */kleta/, dont le français *claie* (mais aussi le francoprovençal *cleya*, l’occitan *clēda*, le gascon *clēda*, etc.) constituent autant de continueurs réguliers. Par ailleurs, la comparaison des langues celtiques permet de supposer un ancêtre commun, dont le gallois */klētā/ est lui-même un représentant. Ces deux démarches convergentes de comparaison-reconstruction, romane et celtique, aboutissent ainsi au même résultat et démontrent que le lexème gaulois a été emprunté à l’époque protoromane par divers parlers de la Romania. Par ailleurs, l’interprétation des données du latin tardif (*clita*, *clida*, etc) permet, en tenant compte de ce que l’on sait des rapports entre les graphies mérovingiennes et la réalisation phonique, de proposer une notation orale de cet ancêtre protoroman: */kleta/. La notice *claie* de TLF-Étym, signée par Nadine Steinfeld⁸, synthétise donc l’analyse nouvelle sous la forme suivante:

Origine. Continueur régulier du protoroman régional */kleta/ subst. fém. “treillis d’osier”, qui se recommande comme ancêtre commun, en plus du lexème français, de francoprovençal *cleya* (depuis 15^e s.), occitan *clēda* (depuis début 13^e s.), gascon *clēda*, piémontais *tsea*, catalan *clēda* (depuis fin 13^e s.), haut aragonais *klēta*, galicien *chedas* (plur.) et portugais *cheda* (depuis 1258) (REW³ 1988; FEW 2, 776a-778b; DECat; KUHN, *Hocharagonesisch* 74; BUSCHMANN; DELP³; HOUAISS). Le type lexical est inconnu du latin écrit de l’Antiquité (Ø TLL), mais on relève son corrélat en latin du haut Moyen Âge du nord de la Gaule: *clita* subst. fém. “treillis” (629/634 [*Lois Ripuaires*]), *clida* “id.” (776 [*Annales regni Francorum*], tous les deux NIERMEYER²). Le lexème est lui-même emprunté à un gaulois */klētā/, dont on trouve d’exacts correspondants dans les langues celtiques: ancien irlandais *cliath* subst. fém. “claie, rangée”, gallois *chwyd* “barrière”, moyen cornique *chuit* “*clēta*”, breton *kloued(enn)* “haie” (VENDRYES, *Lexique*; DOTTIN, *Langue gauloise* 246; LAMBERT, *Langue gauloise*² 195; DELAMARRE, *Langue gauloise*²). Son aire de dispersion permet de supposer qu’il a été diffusé de la Gaule en Ibérie. Cf. VON WARTBURG in FEW 2, 776a-778a, *CLĒTA 1; REW³ 1988; CHAMBON, *MélXYZ* 6.

Alose

TLF (2: 603) fait provenir le français *alose* (attesté pour la première fois dans un texte de la fin du 12^e siècle) du “bas-latin” *alausa*, dénomination mentionnée dès le 4^e siècle dans le poème *Mosella* de l’Aquitain Ausone [Decimus Magnus Ausonius], qui cite ce poisson comme le mets du peuple (“obsonia plebis, alausas”). Pour TLF, le français *alose* est un “emprunt au bas latin”:

⁸ Toutes les notices de TLF-Étym sont relues et amendées par plusieurs relecteurs; la notice *claie* a été relue par Jean-Paul Chauveau, Jean-Pierre Chambon, Gilles Roques, Frankwalt Möhren et Éva Buchi.

Étymol. ET HIST. — Fin XII^e s. ichtyol. (*Roman d'Alexandre*, éd. H. Michelant, 232, 34, ds T.-L.: Autresi com l'alse englotist la balaine).

Empr. au b. lat. *alaua* “genre de poisson” attesté dep. le IV^e s. (AUSONE, *Mos.*, 127 ds *TLL* s.v., 1483, 21: stridentesque focis, obsonia plebis, alausas), d'orig. gaul. (voir A. Thomas ds *Romania*, t. 36, p. 96).

Le FEW (I, 58 et désormais 24/1, 293) avait cependant bien noté que les formes gallo-romanes possédaient des corrélats dans d'autres parlers romans, notamment en catalan, mais il gardait comme étymon le latin *alaua*. La nouvelle notice de TLF-Étym, signée par Rebecca Loeb (avec Éva Buchi, Jean-Pierre Chambon et Andres Kristol comme contributeurs-relecteurs) part des corrélats attestés dans les divers parlers romans (francoprovençal, occitan, italien, catalan) pour reconstruire un ancêtre protoroman commun, la forme française étant elle-même un des représentants parmi d'autres de la protoforme:

Origine. Continuateur régulier du protoroman régional */a'lausa/ subst. fém. “alose”, qui se recommande comme ancêtre commun, en plus du lexème français, de francoprovençal (Le Bourget) *aloušā* subst. fém. “alose” (“poisson du lac du Bourget”, CONSTANTIN, *Savoyard*), occitan *alaua* (attesté depuis 1397 [attesté indirectement depuis 1360, à travers le dérivé *alauat* subst. masc. “filet pour la pêche des aloses”], PANSIER 3), italien *alosa* (depuis 1280/1310, DELI²; LEI 1, 1464), et peut-être catalan *alosa* (“mot rar i mal documentat”, synonyme de *saboga*, DECat). Ce dernier, comme espagnol *alosa* (attesté depuis 1438, ALONSO, *Diccionario medieval*; [“menos corriente que *sábalo*”, DCECH]), pourrait toutefois être emprunté. Le latin *alaua* subst. fém. “*pisciculi genus*” (attesté depuis Ausone [Bordeaux 4^e siècle], *TLL* 1, 1483) confirme cette reconstruction. Très probablement emprunté à une langue de substrat, mais sans corrélats en celtique, le lexème protoroman est d'origine inconnue (il est généralement attribué au gaulois dans la bibliographie, cf. THOMAS, *Romania* 36, 96; REW³ 314; FEW 24, 293b; HUBSCHMID, *ZrP* 66, 56-57; MICHEL, *BALM* 1, 167-177; LAMBERT, *Langue gauloise*² 188; DELAMARRE, *Langue gauloise*² 32). L'aire de dispersion du mot et l'attestation d'Ausone laissent croire en tout cas que sa diffusion s'est produite à partir de la Gaule. Cf. VON WARTBURG/JÄNICKE in FEW 24, 293b, ALAUSA.

Cette reconstruction de la forme orale protoromane */a'lausa/ rencontre, pour ainsi dire, la forme écrite latine *alaua*, mentionnée par l'écrivain Ausone. Mais on voudra bien admettre, d'un point de vue méthodologique, que la forme orale reconstruite ne trouve dans la forme latine qu'une confirmation, et non une légitimation. La reconstruction protoromane, établie sur les corrélats, eût été la même sans le témoignage d'Ausone, elle eût eu simplement moins de certitude. Par rapport au cas de *claire*, cette reconstruction protoromane ne peut s'appuyer sur le deuxième pilier que serait une reconstruction similaire dans une autre aire linguistique. Le manque de corrélats dans les autres parlers celtiques rend en effet l'hypothèse d'une origine gauloise peu probable: on pourrait tout aussi bien, et avec autant de vraisemblance, avancer l'hypothèse d'une origine ligurienne ou d'un parler aquitain. Il est donc plus prudent d'évoquer pour l'heure une langue de substrat inconnue servant d'*etymologia remota* au lexème protoroman.

Bigler

TLF (4: 496-497), à la suite de FEW (1: 380a, s.v. *BISOCULARE) analysait le verbe français *bigler* comme un “emprunt au latin populaire” **bisōcūlare*, cette forme non attestée résultant elle-même d'un verbe supposé **oculare*, formé sur le substantif du latin classique *oculus*, précédé de *bis-*:

Empr. à un lat. pop. **bisōcūlare* “loucher” (*bis* “deux fois, de travers” et **oculare* “regarder” < *oculus*, *œil*); forme *bigler* d'apr. *bigle**. Dans l'hyp., moins vraisemblable, où *bigle*¹* serait préexistant à *bigler* (*EWFS*², v. *bigle*¹), celui-ci en serait dérivé.

La nouvelle version de la notice étymologique de *bigler*, désormais disponible sur le site de l'ATILF, renverse totalement la perspective, et propose une analyse un peu plus complexe.

Renonçant à dériver le mot français directement d'un supposé "latin populaire" reconstruit, les rédacteurs de la nouvelle notice, Jean-Paul Chauveau et Gilles Petrequin, partant des formes attestées en castillan (*bisojo*) et dans deux dialectes gallo-romans (le poitevin et le franc-comtois), reconstituent un étymon protoroman commun aux formes attestées historiquement, étymon présenté en notation phonologique et précédée d'un astérisque, comme il se doit pour une protoforme orale reconstituée: */bi'sOkLU⁹. Mais il fallait encore rendre compte de l'évolution phonétique particulière de *bigler*, qui conserve le groupe /-kl-/ (<protoroman /-skl-/) alors que celui-ci est normalement réduit, dans l'aire oïlique, à /-sl-/ en ancien français puis /-l-/ en français moderne (voir latin *mīscūlare* > afr. *mesler* > frm. *mêler*). L'explication la plus vraisemblable est que la forme *bigler* se soit formée dans la frange méridionale de l'aire oïlique, zone qui connaît un traitement phonétique comparable à celui que l'on retrouve en francoprovençal, où le groupe /-skl-/ donne régulièrement /-kl-/. La nouvelle notice de TLF-Étym résume donc ainsi cette analyse:

Origine. Continuateur régional régulier du protoroman */biso'klare/ v. "loucher", qu'on suppose à partir du protoroman */bi'sOkLU/ adj. "qui louche", représenté, dans une partie de l'aire oïlique, par le poitevin et le franc-comtois *biseuil* "id." (VON WARTBURG in FEW 7, 314a, OCULUS) et par l'espagnol *bisojo* "id." (attesté depuis 1400, DCECH 1, 593). La conservation du /-kl-/ intervocalique oblige à voir dans *bigler* un emprunt par le français, à partir du 16^e siècle, à un dialecte de l'aire oïlique où c'est le traitement phonétique régulier du groupe consonantique /-skl-/. La zone d'origine la plus probable de la diffusion est le sud-est oïlique (Bourgogne, Franche-Comté), où des représentants de *masculu*, *misculāre* et **cremasculu* sont attestés avec ce même traitement, comme en francoprovençal. (VON WARTBURG in FEW 6/1, 427a, MASCULUS; 6/2, 158ab, MISCULARE; 2, 1312a, KREMASTER). Cf. DONDAINE, *parlers comtois*, pages 95-96 et TAVERDET, *Bourgogne*, pages 184-185. Le sud-ouest du domaine oïlique constitue un point de départ moins vraisemblable, puisqu'il ne connaît, par ailleurs, qu'un seul exemple de cette évolution phonétique cf. *MUSCULA>*moucle* subst. fém. "moule, coquillage" (attesté depuis 1500, dans le Sud et l'Ouest du domaine d'oïl, FEW 6/3, 261b-262a, MUSCULUS I 2 a). La diffusion de *bicler/bigler* en français doit être comparable à celle de *racler*, qui relève lui aussi de ce traitement phonétique et qui est d'abord documenté en Bourgogne. (cf. TLF, s.v.*racler*; VON WARTBURG in FEW 10, 79a, *RASCLARE). L'antériorité dans les textes, par rapport à *bigler*, de la forme nominale, *bigle*, employée comme surnom, n'est pas surprenante; elle n'implique pas que ce soit le point de départ de la famille. Cf. VON WARTBURG in FEW 1, 380a, *BISOCULARE. Une autre hypothèse, qui propose le rattachement à *bique* subst. fém. "chèvre" (SPITZER, ZrP 44, 189; SAINEAN, *Étymologie* 1, 68-69; JÄNICKE, *Mél. Deutschmann* 141-159), se heurte à plusieurs difficultés: la forme **bisque* "chèvre" n'est pas attestée et n'apparaît que dans des dérivés tardifs; le passage de **bisque* subst. fém. à *bisclé* adj. serait inattendu; l'intermédiaire le plus vraisemblable devrait être un dérivé verbal en *-eler*: **bisqueler*, qui ne peut pas rendre compte des formes dialectales de type [bi'kle].

Le cas de *bigler* forme, d'une certaine façon, un passage à la limite de la méthode de l'étymologie comparée-reconstruction. L'hypothèse que nous avançons est loin de présenter le même degré de certitude que le cas de *claié*, qui reposait sur une double reconstruction, protoromane et celtique, ou même que celui d'*alose*, qui s'appuyait sur des corrélats répartis sur une aire relativement étendue de la Romania. On pourrait en effet avancer que l'espagnol *bisojo*, récemment attesté, peut résulter d'une formation purement espagnole, de même que le franc-comtois *biseuil* ou le poitevin *bizeuil*, visiblement formés sur le même modèle. Il reste qu'il est peu probable que trois parlers différents, largement dispersés dans la Romania, aient formés indépendamment les uns des autres des formes similaires ayant précisément la même signification. Il est donc de meilleure méthode de considérer que toutes ces formes, de même signification, appartiennent à la même famille étymologique et de postuler un ancêtre commun protoroman qui en rende compte. Par ailleurs, l'hypothèse d'une formation de *bigler* dans les

⁹ Dans */bi'sOkLU/, le dernier caractère, U majuscule, représente la forme prise en protoroman par le ù bref latin, c'est-à-dire un u relâché qui se confondra en protoroman occidental en o fermé.

marges oïliques méridionales permet d'insérer ce mot dans un type de traitement phonétique cohérent, même s'il faut concéder que les documents manquent, à l'heure actuelle, pour localiser très précisément la formation première du mot.¹⁰

L'origine méridionale du mot en français est cependant confirmée par l'aire géolinguistique qu'il couvre dans les parlers dialectaux: la moitié sud du domaine d'oïl (FEW 1: 380), le nord-ouest occitan (ALG 613, ALAL 723*) et le domaine francoprovençal (ALLy 1071, ALJA 1319, GPSR 2, 395). Le caractère autochtone du mot ne fait pas de doute pour des formes comme *biscle* m. "louchon" en Lot-et-Garonne et en Gironde. La répartition dans les parlers dialectaux d'oïl n'est pas due à une importation récente du français, puisque le radical est essentiellement le radical ancien /bicl/- et qu'il a subi le plus souvent la palatalisation du groupe -/kl/- là où celle-ci est régulière. On peut donc conclure que le français *bigler*, tant qu'il demeurerait isolé (par exemple dans l'analyse étymologique du TLF), restait sans explication satisfaisante alors qu'il trouve, rapproché des autres parlers romans, une origine vraisemblable. Par ailleurs, les données de la géolinguistique permettent de délimiter une localisation probable à la fois pour la formation du mot et pour l'origine de sa diffusion en français.

Conclusion

La nouvelle méthodologie mise en œuvre par TLF-Étym concernant le lexique héréditaire français s'est développée parallèlement au renouveau méthodologique qui s'observe dans le domaine de l'étymologie romane. Dans l'un et l'autre cas, il ne s'agit plus de poser un étymon latin, philologiquement attesté, et de voir ce qu'il est devenu dans les parlers romans, mais de partir des données romanes pour reconstruire une forme protoromane de nature orale. Le mouvement de la recherche est donc exactement inverse. Cette reconstruction doit tenir compte —et doit rendre compte dans la mesure du possible— de toutes les données linguistiques disponibles. Il s'agit donc d'intégrer dans l'étymologisation du lexique héréditaire français, à la fois les témoignages attestés dans les autres parlers de la Romania, mais aussi, et sur un pied d'égalité, les corrélatifs fournis par les différents langues, dialectes et patois de l'aire gallo-romane (FEW). Dans le cas des étymons protoromans du lexique français héréditaire, la validité de la reconstruction est naturellement conditionnée par la richesse et par la diversité (tant sur le plan diachronique que sur le plan diatopique) des données linguistiques qui subsistent. La reconstruction de l'étymon atteint, il faut le concéder, un degré variable de probabilité selon le nombre des corrélatifs dont on dispose pour étayer l'hypothèse, la probabilité de véracité étant proportionnelle au nombre des données disponibles. Mais dans tous les cas la méthode reste la même: il convient d'opérer l'intégration, de la manière la plus économique (la plus "élégante" dirait-on en langage mathématique) de toutes les données romanes, en tenant compte à la fois des spécificités de l'histoire interne et de l'histoire externe des témoignages linguistiques conservés.

¹⁰ La première attestation textuelle actuellement disponible se trouve dans un ouvrage de Pierre Belon, *La Nature et diversité des poissons avec leurs pourtraicts representez au plus pres du naturel*, Paris, C. Estienne, publié en 1555: "ils [certains poissons] vont d'un costé bisclant des yeulx". On peut supposer que ce premier exemple est tardif par rapport à une formation orale et populaire.

Références bibliographiques

- [ALAL]. Potte, J.-C. (1975-). J.-C. *Atlas Linguistique et ethnographique de l'Auvergne et du Limousin*. Paris: Éditions du CNRS.
- [ALJA]. Martin, J-B; Tuailon, G. (1971-1978). *Atlas Linguistique et Ethnographique du Jura et des Alpes du Nord*. Paris: Editions du CNRS. 3 volumes.
- [ALG]. Seguy, J. (1967-1974). *Atlas Linguistique et ethnographique de la Gascogne*. Paris: Éditions du CNRS.
- [ALLy]. Gardette, P. *et al.* (1967-1976). *Atlas Linguistique et ethnographique du Lyonnais*. Paris: Éditions du CNRS. 5 volumes.
- Baldinger, K. (1959). "L'étymologie hier et aujourd'hui". *Cahiers de l'Association internationale des Études françaises* 11. 233-264.
- Bourciez, É. (1967⁵ [1910¹]). *Éléments de linguistique romane*. Paris: Klincksieck.
- Büchi, É. (1996). *Les Structures du Französisches Etymologisches Wörterbuch. Recherches métalxicographiques et métalxicologiques*. Tübingen: Niemeyer.
- Buchi, É. (2005). "Le projet TLF-Étym (projet de révision sélective des notices étymologiques du Trésor de la langue française informatisé)". *Estudis romànics* 27. 569-571.
- Buchi, É.; Schweickard, W. (2008). "À la recherche du protoroman: objectifs et méthodes du futur Dictionnaire Étymologique Roman (DÉRom)". Dans Iliescu M. et al. (ed.). *Actes du XXV^e Congrès International de Linguistique et de Philologie Romanes (Innsbruck, 3-8 septembre 2007)*. (à paraître).
- Chambon, J.-P. (2007). "Remarques sur la grammaire comparée-reconstruction en linguistique romane (situation, perspectives)". *Mémoires de la Société de Linguistique de Paris 15 (nouvelle série)*. 57-72.
- Chambon, J.-P. (à paraître). "Pratique étymologique en domaine (gallo)roman et grammaire comparée-reconstruction. À propos du traitement des mots héréditaires dans le TLF et le FEW".
- [DMF]. *Dictionnaire du Moyen français*. Action de recherche de l'équipe Linguistique historique française et romane de l'ATILF.
- [DMF2]. Martin, R. (dir.). *Dictionnaire du moyen français (DMF) (1330-1500)*. Seconde version: DMF2, Nancy, 2007. Base de données informatisée consultable en ligne <http://www.atilf.fr/dmf> [Date de consultation: mars 2008].
- Dottin, G. (1920). *La Langue gauloise. Grammaire, textes et glossaire*. Paris: Klincksieck.
- [FEW]. Wartburg, W. von et al. (1922-2002). *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes*. 25 volumes. Bonn/Heidelberg/Leipzig-Berlin/Bâle: Klopp/Winter/Teubner/Zbinden.
- Fischer, I. (1969). "Lexicul. Fondul panromanic. L". Dans Coteanu, I. (ed). *Istoria limbii române 2*. Bucarest: Editura Academiei Republicii socialiste România. 110-116.
- [GPSR]. Gauchat, L.; Jeanjaquet, J.; Tappolet, E. et al. (1924-). *Glossaire des patois de la Suisse romande*. Neuchâtel, Paris, Genève.
- Hall, R. A. Jr. (1950). "The Reconstruction of Proto-Romance". *Language*. 266-27.
- Martin, R. (1998). *Dictionnaire du Moyen Français, fascicule A-AH*. Nancy: CNRS-InaLF.
- Niermeyer, J. F. et al. (2002² [1954-1976¹]). *Mediae Latinitatis lexicon minus: lexique latin médiéval*. Leiden: Brill. 2 volumes.
- [REW]. Meyer-Lübke, W. (1935³ [1911-1920¹]). *Romanisches Etymologisches Wörterbuch*. Heidelberg: Winter.
- [TLF]. Imbs, P.; Quemada, B. (dir.) (1971-1994). *Trésor de la Langue Française. Dictionnaire de la langue du XIX^e et du XX^e siècle (1789-1960)*. Paris: Éditions du CNRS/Gallimard. 16 volumes.
- [TLFi]. Version informatisée du TLF et consultable en ligne à l'adresse <http://atilf.atilf.fr/>. [Date de consultation: en mars 2008].
- [TLF-Étym]. Révision sélective des notices étymologiques du Trésor de la langue française informatisé [en ligne]. <http://www.atilf.fr/tlf-etym> [Date de consultation: mars 2008].