

Myriam Benarroch (Université Paris-Sorbonne)

L'apport du DÉRom à l'étymologie portugaise

Le *Dictionnaire Étymologique Roman* (DÉRom) est le fruit d'un projet franco-allemand inauguré lors du précédent CILPR, à Innsbruck, en septembre 2007 (cf. Buchi / Schweickard 2010). L'équipe du DÉRom, dirigée par Éva Buchi et Wolfgang Schweickard, est composée de 45 linguistes romanistes, d'un informaticien et de deux responsables de la documentation. Le dictionnaire, publié sur internet, compte à ce jour 33 articles et plus d'une centaine sont en cours de rédaction. Nous nous appuyerons sur les articles publiés ainsi que sur les nombreuses discussions méthodologiques qui ont animé rédacteurs et réviseurs pour tenter de mettre en lumière ce que le DÉRom, par sa conception renouvelée de l'étymologie héréditaire romane, peut apporter à l'étymologie portugaise.

L'objectif du DÉRom est de reconsidérer l'étymologie du lexique héréditaire commun à l'ensemble des idiomes romans à la lumière des avancées de la linguistique historique romane. L'utilisation, en particulier, de la méthode de la grammaire comparée-reconstruction (cf. Chambon 2007; 2010; de Dardel 2007; 2009) permet, à partir des cognats («lexèmes reliés par un réseau de correspondances phoniques et sémiques», Buchi 2010: 44) de reconstruire les étymons protoromans qui constitueront les lemmes du DÉRom: ils sont présentés en notation phonologique et portent l'astérisque, ex. */ka'βall-u/.

La métalangue du DÉRom est le français. Les normes rédactionnelles, la bibliographie ainsi que d'autres outils nécessaires à la rédaction et à la révision des articles sont réunis dans un ouvrage mis à jour deux fois par an et intitulé *Livre bleu* (LB). Les articles du DÉRom comportent obligatoirement les parties suivantes: (1) «Lemme»; (2) «Matériaux»; (3) «Commentaire»; (4) «Bibliographie»; (5) «Signatures»; (6) «Date de mise en ligne de cet article». Selon les normes rédactionnelles, vingt idiomes doivent être obligatoirement cités s'ils possèdent un continuateur du protoroman. Pour les idiomes romans de la péninsule Ibérique, le DÉRom considère cinq idiomes «obligatoires»: catalan, espagnol, asturien, galicien et portugais. Sur les 33 articles publiés, seuls quatre lexèmes protoromans (*/a'pril i-u/; */βi'n-aki-a/; */ka'βall-a/; */karpIn-u/) n'ont pas de continuateur en portugais. Notre travail portera donc sur les 29 articles restants.¹

¹ */a'gʊst-u/; */'ali-u/; */'anIm-a/; */'ann-u/; */a'pril-e/; */as'kʊlt-a /; */'aud-i-/; */'baβ-a/; */'bɪβ-e/; */'βIndIk-a-/; */'dɛke/; */'ɛder-a/; */'ɛrβ-a/ ~ */'ɛrb-a/; */; */es'kʊlt-a-/; */'eks-i/; */'Φak-e-/; */'Φe'βr-ar-i-u/; */'Φɛn-u/ ~ */'Φɛn-u/; */ka'βall-u/; */'kad-e-/; */'karn-e/; */kas'tani-a/ ~ */kas'tini-a/; */ka'ten-a/; */'kul-u/; */'laks-a-/; */'mai-u/; */'mart-i-u/; */'mɔnt-e/; */'pɔnt-e/.

1. Quelques problèmes généraux de l'étymologie portugaise

1.1. Les sources

Le premier problème posé à l'étymologie portugaise réside dans le retard pris dans la connaissance, le dépouillement et l'analyse des textes médiévaux. De là, le faible nombre d'éditions critiques et l'absence d'un corpus conséquent et facilement accessible. Toutefois, les choses avancent à grands pas: plus de trente textes rédigés entre 1175 et 1255 ont été découverts ces dernières années (cf. Martins 2007). Le dictionnaire étymologique de référence est toujours le ³DELP de José Pedro Machado en cinq volumes, qui a plus de trente ans (1977). Cette œuvre d'un seul homme a le double mérite d'être la plus complète du point de vue de la nomenclature et de donner –et elle est la seule à le faire– les premières attestations accompagnées de la citation et de la source. Toutefois, elle ne possède pas d'introduction, ne dit rien de la méthodologie utilisée et montre une bibliographie lacunaire et non exploitée systématiquement. Les travaux d'Antônio Geraldo da Cunha (1982; 1986-1994; 1992; ²2006/2007 [¹2002]) représentent un saut qualitatif considérable par rapport au ³DELP. Le *Houaiss* (2001), tout en étant un dictionnaire de langue, constitue à ce jour le dictionnaire étymologique et historique du portugais le plus à jour, en particulier grâce à l'exploitation qu'il fait des travaux de Cunha. Il présente toutefois de graves défauts méthodologiques (cf. Benarroch 2010), en particulier une non exploitation systématique des sources citées et un manque de précision dans la citation des sources rendant difficile, voire impossible, leur vérification.

Le DÉRom comporte une bibliographie «de consultation et de citation obligatoires», classée par domaines géographiques. Pour le portugais, elle réunit les ouvrages suivants: Cunha (1986-1994; 1992; ²2006/2007 [¹2002]); ³DELP; *Houaiss*; Lisboa (1937); Piel (1932); ALPI.² À ces sources portugaises (et brésiliennes), viennent s'ajouter les sources galiciennes, indispensables pour rédiger la partie de l'article consacrée au portugais: DDGM; Buschmann (1965); DRAG. Pour être publié, tout article du DÉRom doit avoir été relu par au moins un des réviseurs par domaine géographique. La révision des sources citées par les dictionnaires portugais permet de s'assurer qu'elles sont bien rédigées en portugais et non en latin (cf. *infra* 5.1) et de corriger des erreurs dans les citations de textes³ ainsi que dans les références aux sources: ex. *s.v. feno* dans ³DELP (cf. *infra* note 6).

1.2. La langue des textes médiévaux: galicien, portugais et galégo-portugais

La réflexion sur la langue parlée dans le Nord-Ouest de la péninsule Ibérique est parasitée, tant en Galice qu'au Portugal, par des positionnements idéologiques bien éloignés de questions purement linguistiques. Il n'existe pas de consensus sur la question (cf. Martins 2007: 161-162). On peut toutefois affirmer que le galicien et le portugais sont issus d'une même langue orale médiévale –dialecte primaire issu du latin parlé–,

² Navarro Tomás, Tomás (dir.) (1962): *Atlas Lingüístico de la Península Ibérica*. Madrid: CSIC.

³ Par ex., *s.v. fevereiro* in ³DELP, «... et de odie die uel tempore quod erit VX (sic) kalendas feuerarias et era millesima XXIIIa...», il faut corriger *et* en *ut* et *feuerarias* en *feueruarias* (cf. Herculano 1867: 90).

que sur une rive du Minho l'on appelle *galego*, sur l'autre, *galego-português*. L'emploi même du glottonyme utilisé pour désigner cette langue médiévale n'est pas neutre, ni simple. Au Portugal, la tendance actuelle est de réserver le nom de *galego-português* à la langue littéraire des *cantigas* et de désigner la langue orale ou celle des textes notariaux et juridiques par *português antigo* (cf. Castro 2008).

Cette question du glottonyme s'est posée dès le premier article rédigé dans le DÉRom et a fait l'objet de nombreuses discussions. Les solutions adoptées ont évolué au fur et à mesure qu'apparaissaient de nouveaux problèmes. Celle qui s'est stabilisée depuis quelque temps, mais qui n'est peut-être pas définitive, est la suivante:

- (1) Quand un lexème ou un grammème commun au galicien et au portugais est attesté dès la période galégo-portugaise (avant le milieu du 14^e siècle), il porte le glottonyme «**gal./port.**» et est pourvu d'une seule datation. Les informations se présentent alors sous la forme: «**gal./port.**» + lexème + datation + références bibliographiques (d'abord celle qui fournit la première datation, puis celles concernant le galicien et/ou les deux idiomes [...], enfin celles concernant le seul portugais) [...].
- (2) Si les signifiants du cognat galicien et portugais sont différents, la présentation est de la forme «**gal.**» + cognat galicien + «/» + «**port.**» + cognat portugais + datation + références bibliographiques [...]
- (3) Si le galicien et le portugais ont eu au Moyen Âge un représentant héréditaire commun, mais que ce dernier a été évincé en galicien ou en portugais contemporain par un lexème ou un grammème non héréditaire, les deux idiomes sont distingués sous les glottonymes respectifs «**gal.**» et «**aport.**» (ancien portugais) ou bien «**agal.**» (ancien galicien) et «**port.**». Seule la forme héréditaire figure dans les matériaux, la forme non héréditaire contemporaine étant signalée dans une note (LB: 50).

1.3 La question du latin

L'autre question qui agite les historiens de la langue portugaise est celle de savoir en quelle langue sont écrits les «premiers» textes médiévaux (cf. Castro 2004; Emiliano 2003; Lorenzo 2007; Martins 1999; 2007; Souto Cabo 2004). Avec deux aspects essentiels: la datation et la nature de la langue. En 1901 déjà, Leite de Vasconcelos (1987: 14) avait mis en garde contre la manière d'utiliser les documents «latino-barbares». Ivo Castro (2004: 78), à son tour, affiche la plus grande prudence quant à la désignation et au statut linguistique des textes rédigés entre 882 et 1255.

Comment les dictionnaires étymologiques portugais envisagent-ils les textes latins? Le ³DELP, la plupart du temps repris par *Houaiss*, considère comme première attestation d'un lexème portugais sa première apparition dans un texte, qu'il soit en portugais ou en latin. Les plus anciens documents qu'il cite ont été rédigés entre le 9^e et les 12^e-13^e siècles, ce qui correspond à la période évoquée par Ivo Castro et que Leite de Vasconcelos (1987: 12) qualifie de *protohistorique*. Prenons quelques exemples parmi nos 33 lexèmes du DÉRom. Pour *abril*, *agosto* et *fazer*, ³DELP ne fournit pas de première attestation dans un texte

portugais.⁴ Dans le cas de *cair*, il ne propose une première attestation (au 15^e s.) que pour la forme contemporaine; pour la forme héréditaire, *caer*, donnée comme «arc[aïsme]» avec la date de 1152, il ne fournit pas la citation qui, vérification faite, est en latin. *Houaiss* reprend pour l'essentiel (sauf dans le cas de *cair* où il se réfère à Cunha 1988) les dates et sources de ³DELP, sans plus d'interrogation sur la langue des textes cités, et il inclut dans les «[formes] hist[oriques]» des formes plus latines que portugaises, telle *cauallus*, forme à laquelle il attribue la première attestation (en 870).

Qu'en est-il du latin dans le DÉRom? Faire le choix de présenter comme lemme de l'article du DÉRom non un lexème latin mais la forme phonologique protoromane reconstruite relève d'une démarche visant à privilégier la langue parlée par rapport à la langue écrite et caractérisée par le fait que l'étymon ainsi reconstitué n'est pas le point de départ de l'analyse étymologique mais l'aboutissement du travail de reconstruction (cf. Chauveau 2010: 1). Cela ne signifie aucunement que le latin écrit de l'Antiquité n'a pas sa place dans ce dictionnaire: la forme latine est toujours citée, lorsqu'elle existe, dans la partie «Commentaire», comme corrélat des issues romanes. Tout article du DÉRom peut donc être interrogé à partir de son corrélat latin ou de l'entrée correspondante du REW. Toutefois, si l'on trouve des attestations latines remontant au 3^e s. av. J.C. – *annus* est attesté depuis Naevius) –, le latin écrit de l'Antiquité ne connaît pas nécessairement de corrélat: s.v. */es'kʊlt-a/, «Le latin écrit de l'Antiquité ne connaît pas de corrélat de ce lexème, mais des attestations du type *abscultare* (fin 2^e s. – 6^e s.) et *obscultare* (45/43; 6^e s.[...]) témoignent de la réanalyse en un verbe préfixé de *auscultare* effectuée par les locuteurs». Le corrélat n'existe parfois que pour un seul sens quand le protoroman connaît deux ou plusieurs sémantismes: ainsi, */anIm a/ n'a de corrélat en latin écrit de l'Antiquité que pour le sens abstrait «partie immatérielle des êtres, âme», le second sens, concret, «organe central de l'appareil circulatoire, cœur; partie renflée du tube digestif, estomac», ne s'étant développé qu'en protoroman puis en roumain et en sarde.

2. Le signifiant de l'étymon

2.1. Le signifiant de l'étymon dans les dictionnaires étymologiques portugais

Chacun des trois dictionnaires que nous avons pris pour référence a sa manière de présenter la forme de l'étymon latin. *Houaiss* fournit systématiquement la forme du latin écrit de l'Antiquité: pour les substantifs, par exemple, le nominatif (*cabāllus*) et, au besoin, le génitif, (*caro*, *carnis*). ³DELP donne systématiquement la forme de l'accusatif apocopée (*ponte*–; *carne*–). DENF n'est pas systématique: *caballus*; *cūlus*, -i; *majūs*; *martius*; *pōns pontis* mais *annum*; *aprīlem*; *dēcēm*. Dans la plupart des cas, le lexème est donné comme provenant du «latin» sans autre spécification. Peu d'étymons ont droit à une précision diastatique ou diamésique: *alho* («Do lat. *aleu*–, «alho», considerada forma vulgar por Porfirio», ³DELP; «Do lat. *al(l)ium*, de *alum*», DENF; «lat. *allium*, *īi* «alho; cebola», pelo vulgar», *Houaiss*). Lorsque précision il y a, elle figure rarement dans les trois dictionnaires: DENF est le seul

⁴ Pour *abril*, il mentionne toutefois un anthroponyme.

à donner «lat[im] vulg[ar] *agŭstus*» comme étymon de *agosto*; ³DELP, le seul à mentionner «lat[im] pop[ular] **cadēre*» pour *cair*, le seul aussi à mentionner une réfection savante pour *beber* «com possível recomposição culta, pois a forma ant. era *bever*». Houaiss fournit, quoique rarement, des informations complémentaires: s.v. *ouvir*: «através de uma f[orma] **ouir*»; s.v. *vingar*: «f[orma] div[er]g[ente] vulg[ar] de *vindicar*». S.v. *escutar*, l'étymon proposé, «lat. *ausculto*, *as*, *āvi*, *ātum*, *āre*» est opposé à *auscultar* «cp. *auscultar*, f[orma] div[er]g[ente] erud[ita]»; *feno* porte la mention «lat[im] cl[ássico]», sans qu'il soit dit pour autant que *fêo*, pourtant mentionné, est la forme héréditaire.⁵ Dans les trois dictionnaires, l'étymon de *baba* est donné sous une forme précédée d'un astérisque (**baba*) tout en étant accompagné du même glottonyme passe-partout «lat[im]».

Ces exemples révèlent dans nos dictionnaires une non prise en compte de la variation diamésique à l'intérieur du diasystème latin et, au mieux, une faible prise en compte du latin parlé, qu'il s'agisse du latin de l'Antiquité ou du latin médiéval.

2.2. L'étymon n'est pas toujours celui qu'on croit

Dans l'article *escutar* de ³DELP, l'étymon proposé est «lat. *auscultāre*», tant pour les formes historiques *ascoitar*, *ascuitar* (13^e s.), que pour *escoitar*, *escuitar* (respectivement 13^e et 14^e s.). La recherche de l'étymon protoroman entreprise dans le DÉRom à travers la comparaison entre toutes les issues romanes montre que les formes *ascoitar*, *ascuitar*, d'une part et *escoitar*, *escuitar*, d'autre part, remontent à deux bases étymologiques distinctes, */as'kult-a/ et */es'kult-a/. DÉRom consacre donc deux articles distincts à chacun de ces lemmes protoromans.

Mais il arrive aussi que, si l'on peut reconstruire un étymon protoroman commun à l'ensemble de la Romania, se dégagent des sous-ensembles, souvent aréologiques, remontant, pour des raisons phonétiques, morphologiques ou sémantiques, à un étymon intermédiaire. Dans ce cas, l'article de DÉRom est subdivisé selon les différents types.

2.3. Subdivisions des articles du DÉRom par types phonétiques

Parmi les 29 lexèmes ayant des continuateurs réguliers en portugais, deux ne sont pas arrivés jusqu'à nos jours. L'un, *fêo* /*feo*, 1260/1284 – 15^e s., a été évincé par un emprunt savant au latin, *feno*. L'autre, *caer*, attesté jusqu'en 1452/1453, a subi des changements par analogie et il est entré en concurrence avec *cair*, attesté dès 1259, qui l'a peu à peu remplacé. 6 des 29 articles où le portugais est représenté comportent des subdivisions par types phonétiques: (1) */ɛder-a/: I. */ɛder-a/, II. */ɛler-a/, III. */ɛlen-a/; (2) */ɛrb-a/ ~ */ɛrβ-a/: I. */ɛrb-a/, II. */ɛrβ-a/; (3) */ϕak-e-/: I. Sans accident phonétique */ϕak-e-re/, II. Avec accident phonétique */ϕ-a-re/; (4) */ϕen-u/ ~ */ϕɛn-u/: I. */ϕen-u/ ~ II. */ϕɛn-u/; (5) */kas'tani-a/ ~ */kas'tɪni-a/: I. */kas'tani-a/, II. */kas'tɪni-a/; (6) */laks-a-/: I. */laks-a-/, II. */laks-i-a-/.

⁵ Contrairement à ³DELP: «Trata-se, evidentemente, de reconstrução culta; pois, por via normal, o cit. voc. lat. deu *fêo*, *feo* no séc. XIII, segundo L. V. (Opúsculos, I, p. 259, em excelente artigo consagrado especialmente a este voc.» (N.B. Il faut lire «p. 527-529» et non «p. 259»).

Hormis le cas (2) toutes les issues portugaises relèvent du type I qui est le plus ancien et/ou le plus répandu. Dans les cas (2) et (5), on ne peut reconstruire plus haut pour trouver un ancêtre commun à l'ensemble des issues romanes: «**/ɛrb-a/* et **/ɛrβ-a/* sont irréductibles en reconstruction et donc à considérer comme des variantes protoromanes». C'est pourquoi le lemme est double. La fricative du portugais montre que l'issue portugaise remonte au type II. Si le port. *fazer* (3) relève du type I., majoritaire, le type II. se retrouve dans la conjugaison du futur *farei* et du conditionnel *faria*. Comme il est dit dans le commentaire: «les deux types présentent une répartition variée au sein de leurs paradigmes flexionnels, ce qui témoigne de leur monogenèse». C'est pourquoi, contrairement à la situation précédente, nous avons un seul lemme protoroman, dans un article comportant deux subdivisions, correspondant aux deux types d'infinitif sur lesquels reposent les deux issues romanes.

3. Le traitement des changements morphologiques

Les changements morphologiques sont peu commentés dans les dictionnaires étymologiques portugais. Dans le DÉRom, l'analyse de ces changements, lorsqu'ils se sont produits en protoroman, affecte la structure même de l'article, qui fait alors l'objet de subdivisions par types morphologiques. Trois articles présentent ce type de subdivisions: (1) **/kad-e-/*, (2) **/pɔnt-e/* et (3) **/karpIn-u/*. Le premier reflète un changement de conjugaison, les deux autres, un changement de genre. L'article **/kad-e-/* connaît deux subdivisions: I. 3^e conjugaison: **/kad-e-re/*, II. 2^e conjugaison **/ka'd-e-re/*. L'ancien portugais *caer* relève du second prototype: le changement de conjugaison *-er > -ir* ayant abouti à l'actuel *cair* ne remonte pas au-delà de l'époque médiévale. Voilà le type d'informations que l'on ne trouve pas dans les dictionnaires portugais. L'article **/pɔnt-e/* illustre bien le changement de genre: I. Substantif masculin originel: sarde; II. Substantif féminin: aires latérales et aires isolées; III. Substantif masculin innovant: Romania centrale. D'ou: «[...] protorom. **/pɔnt-e/* connaissait les deux genres, le masculin étant plus ancien, le féminin –issu plus récemment (mais assez tôt pour avoir pu être transmis au roumain) de la tendance analogique à féminiser les substantifs de la troisième déclinaison –étant devenu hégémonique avant d'être repoussé par le masculin innovant».

Pour le portugais, ³DELP et *Houaiss* se contentent de mentionner le changement de genre, ils n'expliquent pas pourquoi *ponte* est féminin en portugais quand il était masculin en latin. L'article **/pɔnt-e/* du DÉRom montre que ce lexème a connu trois états distincts correspondant à des époques différentes et à diverses régions de la Romania, le portugais féminin se situant dans l'étape intermédiaire propre aux aires latérales et aux aires isolées. L'article **/karpIn-u/* compte deux subdivisions morphologiques: I. changement de genre: **/karpIn-u/* s.m.; II. Remorphologisation: **/karpIn-a/* s.f. Le portugais n'est pas représenté dans les matériaux car il ne connaît pas d'issue de cet étymon. Nous devons, par conséquent, corriger les articles *carpa* de ³DELP et de *Houaiss*, le premier donnant ce lexème comme issu d'un masculin *carpinŭ*, le second, d'un féminin précédé de l'astérisque **carpina*, **carpea*. L'article du DÉRom démontre que, en vertu de «considérations d'ordre phonétique, chronologique et géobotanique (le

charme n'est pas autochtone sur la péninsule Ibérique)», tant en portugais qu'en espagnol, ces lexèmes ne sont pas héréditaires, et il conclut, suivant en cela DCECH, qu'il s'agit d'emprunts, apparemment à l'occitan. Nous voyons combien la comparaison intraromane est utile à l'étymologie de chaque idiome roman, ici, à celle du portugais.

4. L'aspect sémantique

Les dictionnaires étymologiques portugais sont très économes en matière de traitement sémantique. A l'exception de *mājus*, le DENF ne donne pas de signifié de l'étymon latin; il propose, en revanche, une définition du lexème étymologisé. ³DELP fournit le sens du latin mais pas celui du portugais. Dans les deux cas, l'absence de signifié d'un des deux idiomes incite à penser qu'il est identique dans les deux, ce qui, du reste est clairement suggéré dans des signifiés réduits à «mesmo sentido» (³DELP, s.v. *abril*; *março*; *ponte*) ou «id.» (*Houaiss*, s.v. *abril*, *março*). Or de nombreux cas échappent à cette équivalence. Quand il est donné, le signifié de l'étymon est généralement rendu par la traduction en portugais du lexème latin et de ses différents sémantismes: «*anīma* <sopro, ar; alento; o princípio da vida; a alma, por oposição ao corpo>», ³DELP; «*anīma,ae* <sopro, ar; alento; o princípio da vida; a alma, por oposição ao corpo>», *Houaiss*, qui a visiblement repris la définition de ³DELP). *Houaiss* fournit, mais uniquement dans 3 cas sur 29 (*bibēre*, *annus*, *maius*), une définition componentielle, avant de donner la traduction du latin. Rien n'est dit, dans ces trois dictionnaires, sur la manière dont on passe du signifié latin à celui du portugais.

Le DÉRom donne systématiquement et uniquement une définition componentielle pour le lemme. Il évite tout item transductif: «*/dēke/ num. card. <neuf plus un>»; «*/ēks-i / v.intr. <aller hors d'un lieu>». En revanche, la traduction en français est proposée dans les matériaux pour le premier idiome cité (le dacoroumain), suite à la définition componentielle: «*/dēke/ > dacoroum. *zece* num. card. <neuf plus un, dix>»; «*/ēk's-i-re/ > dacoroum. *ieși* v.intr. <aller hors d'un lieu, sortir>». En effet, la reconstruction ne s'effectue pas seulement sur le signifiant: à partir des parlers romans, plusieurs sémantismes peuvent être reconstruits. Dans ce cas, ils apparaissent dans la définition du lemme: «*/βIndIk a / v.tr. <faire échapper (quelqu'un) à un danger>; <dédommager moralement (quelqu'un) en punissant (son) offenseur>» et donnent lieu à des subdivisions de l'article.

Les subdivisions sémantiques ne se justifient que si les changements de sens se produisent en protoroman. Quatre articles comportent ce type de subdivisions: (1) */anIm a/: I. sens abstrait = «partie immatérielle des êtres, âme», II. sens concret = «organe central de l'appareil circulatoire, cœur; partie renflée du tube digestif, estomac»; (2) */as'kUlt-a-/: I. «tendre l'oreille vers ce qu'on peut entendre», II. «accueillir avec faveur (les paroles de quelqu'un)»; (3) */es'kUlt a-/, *id.*; (4) */βIndIk-a-/: I. «sauver» > «guérir», II. «venger». Dans le cas de */βIndIk-a-/, ce bisémisme existe dans le corrélat du latin de l'Antiquité. («libérer, sauver» (dp. Plaute [* ca 254 – †184]) et «venger» (dp. Caton l'Ancien [* 234 – †149])). Or la définition du latin *vindico* dans *Houaiss* ne tient pas compte du sens I. Pour */anIm-a/, le latin écrit de l'Antiquité ne connaît pas de corrélat de II.

5. L'apport du DÉRom en matière de datation

5.1. La datation dans les dictionnaires portugais

Nous nous proposons de comparer ici les datations du DÉRom avec celles fournies par ³DELP et *Houaiss*. Nous considérons comme apport du DÉRom toute donnée nouvelle non prise en compte par au moins un de ces deux dictionnaires, sachant que le plus à jour est généralement *Houaiss*. Nous laissons de côté DENF qui n'apporte rien en matière de datation par rapport à *Houaiss*.

Pour les lexèmes du DÉRom que nous avons étudiés, les deux sources principales utilisées par *Houaiss* pour la datation sont ³DELP et les travaux d'Antônio Geraldo da Cunha. Les auteurs du *Houaiss*, publié en 2001 avant la parution du cédérom (Cunha 2002), ont pu, en effet, bénéficier du fichier manuel de Cunha (qu'il désigne sous le sigle «FichIVPM») conservé à la Fondation Casa de Rui Barbosa à Rio de Janeiro, à l'origine du cédérom, publié après la mort de l'auteur. Ce cédérom qui reprend et complète les quatre fascicules de Cunha 1986-1994 est incomplet, certaines données de ces fascicules n'ayant pas été prises en compte.⁶ Ainsi, *alho* et *cadeia* présentent dans cet ouvrage des dates antérieures à celles du cédérom, respectivement 1254 vs 1269 et 1265 vs 13^e s., et *baba* est absent du cédérom. D'autre part, si ²Cunha Vocabulário ne prend en compte que des textes portugais rédigés entre 1200 et 1500, ³DELP donne souvent comme date de première attestation d'un lexème celle d'un texte en latin. À l'opposé, la première date mentionnée dans ³DELP correspond parfois à celle de la forme contemporaine, les formes médiévales étant indiquées comme variantes, suivies de leurs dates respectives (cf. s.v. *escutar*). *Houaiss* ne se préoccupe pas non plus de savoir si le document de première attestation est portugais ou latin et se contente le plus souvent de reprendre la date fournie par ³DELP: «*cavalo* s.m. (870 cf. JM³) [...] etim lat. *cabāllus*, i «cavalo castrado, esp. cavalo de trabalho»; ver *caval*;- f.hist. 870 *cauallus*, sXIII *cavalo*, sXIII *cauallo*, sXV *quaualo*». Or il est parfois difficile de savoir si, dans un texte rédigé en latin, le lexème à étymologiser est encore latin où déjà portugais. Ainsi, ³DELP qui ne date pas port. *ponte*, se contente, pour illustrer le changement de genre de ce lexème, de l'affirmation suivante: «o facto é que no séc. IX já se dizia *illa ponte*». Enfin, d'une manière générale, les dictionnaires portugais ne prennent pas assez en compte les données de la documentation galicienne, pourtant précieuses comme nous allons le voir.

5.2. Les antédations du DÉRom

Dans le domaine de la datation, l'apport de DÉRom est particulièrement important en ce qu'il modifie très souvent les dates de première attestation proposées par les dictionnaires étymologiques, en les avançant ou en les reculant. La réunion et la confrontation des ouvrages de la «bibliographie de consultation et de citation obligatoires» du DÉRom permet de mettre

⁶ Monjour (2004 : 149-150) avait déjà relevé que pour 1,5% des entrées de l'échantillon étudié, la date de première attestation donnée dans *Houaiss* est postérieure à celle des dictionnaires précédents, non qu'il s'agisse d'une postdatation argumentée mais parce que l'auteur semble ne pas avoir pris en compte l'ensemble des données de Machado ou/et de Cunha.

à jour et d'harmoniser les datations suggérées par l'une ou l'autre de ces sources. Parmi les sources consultées pour le sous-domaine portugais, sont systématiquement pris en compte les ouvrages de référence du galicien, les plus importantes d'entre elles étant le DDGM (*Dicionario de Dicionarios do Galego Medieval*) et le TMILG (*Tesouro Medieval Informatizado da Lingua Galega*), qui citent des sources galiciennes et portugaises et sont tous deux accessibles sur internet.

Nous considérons comme antédation pour DÉRom toute date d'un lexème portugais, rédigé dans un texte médiéval galicien/portugais (galicien ou / et portugais), antérieure aux dates proposées par ³DELP et par Houaiss, ou plus précise que ces dernières, à condition que ces dates correspondent à des textes galiciens/portugais.

Nous avons pu antédater 16 des 29 lexèmes étudiés ou en donner une date plus précise, sachant que nous ne prenons pas en considération les dates proposées par ces dictionnaires lorsqu'elles correspondent à des textes en latin. Si le lexème est attesté de manière continue et encore présent dans le portugais contemporain, la date est précédée de *dp.* «depuis»; dans le cas contraire, le glottonyme porte la marque *a-* «ancien» et la date de première attestation est suivie de la date de dernière attestation (cf. *infra*); lorsque le galicien et le portugais présentent deux formes graphiques divergentes ou que la forme est différente de celle du portugais contemporain, elle est mentionnée entre crochets. Nous indiquons ici les lexèmes antédatisés, suivis de la date du DÉRom vs la plus ancienne des deux dates proposées par ³DELP et/ou Houaiss: *alma* (dp. 1214 vs 13^e s.); *ano* (dp. 1214 vs 1264); *aport. ascoitar, ascuitar* (1220/1240 – 14^e s. vs 13^e s.); *baba* (dp. 1295/1312 [*baua*] vs 14^e s.); *carne* (dp. av. 1282 vs. 13^e s.); *castanha* (dp. 1209 [*castañas*] vs 1269); *cavalo* (dp. 1273 vs ø); *cu* (dp. 1220/1240 [*cuu*] s. vs 14^e s.); *erva* (dp. 1240/1300 vs 13^e s.); *escutar* (dp. 1240/1300 vs 13^e s.); *fazer* (dp. 1170/1220 vs 13^e s.); *aport. fêo* (1264/1284 – 15^e s. vs 13^e s.); *fevereiro* (1253/1254 vs 1256); *aport. leixar* (1220/1240 – 1552 vs 13^e s.); *março* (dp. 1253/1254 vs 1270); *ponte* (dp. 1254 vs 13^e s.).

Nous constatons que Houaiss utilise comme principale source pour la datation les travaux de Cunha (11 lexèmes sur 29) et secondairement le ³DELP, plus ancien (4 lexèmes). Les 16 lexèmes ici antédatisés l'ont été grâce aux deux sources galiciennes mentionnées plus haut: DDGM (9) et TMILG (7). Ces sources donnent des dates de première attestation y compris pour des textes on ne peut plus portugais: la date de 1214 que nous avons proposée pour *alma* et *ano* correspond, en effet, à celle de la rédaction du testament d'Alphonse II, roi du Portugal. Ces deux lexèmes sont déjà ainsi datés par Ramón Lorenzo (1968) qui corrigeait la première édition du DELP. Ces corrections n'ont été prises en compte ni par les éditions postérieures du DELP, ni par les autres dictionnaires portugais. Dans le cas de *fazer*, la datation proposée dans les dictionnaires portugais (991) correspondant à un texte latin, celle du DÉRom (1170/1220) représente la date de première attestation. Le DÉRom permet en outre, quoiqu'encore de façon marginale étant donné le peu de lexèmes concernés, de proposer une date pour des sémantismes. Ainsi, nous pouvons dater les deux sens de *aport. ascoitar* (I. 13^e– 14^e s.; II. 1220/1240) et *escutar* (I. 1240/1300; II. 15^e s. [*escuitar*]).

5.3. Les postdatations du DÉRom

Nous avons vu qu'il n'existe pas de consensus sur les documents considérés comme les plus anciens textes de la langue portugaise. Il ressort néanmoins de la confrontation entre spécialistes

des premiers textes portugais (1) qu'aucun d'entre eux ne considère comme portugais des textes antérieurs à la deuxième moitié du 12^e siècle et (2) que tous pensent que parmi les textes étudiés à ce jour, les plus anciens ne remontent pas au-delà de 1175 (1173 pour Emiliano 1999; 2003). Nous considérons donc que, conformément au principe établi dans DÉRom selon lequel les unités romanes ne sont pas datées «à travers une attestation d'apparence romane relevée dans un texte alloglote (notamment latin ou slave)» (LB: 42), les textes antérieurs à 1175 ne sont pas à prendre en considération pour la datation de lexèmes portugais. Nous avons, malgré tout, vérifié que les dates proposées par ³DELP et *Houaiss* que nous avons reculées correspondent toutes à des textes latins ou, en tout cas, à des textes où prédomine le latin. Nous considérons donc comme postdatation pour DÉRom toute date d'un lexème portugais rédigé dans un texte médiéval galicien/portugais pour lequel ³DELP et *Houaiss* donnent une date correspondant à un texte antérieur à 1175 ou à un texte clairement identifié comme latin ou à dominante latine. Toutefois, dans le cas où *Houaiss* et/ou ³DELP fournissent deux dates, l'une d'un texte en latin, l'autre d'un texte en portugais, nous considérons qu'il n'y a pas postdatation et ce, bien que *Houaiss* mentionne la première dans la rubrique «datation» et la considère donc comme première attestation (cf. *supra* 1.3 *cavalo* dont la datation donnée, 870, correspond à la forme *cauallus*). Le nombre des postdatations s'en trouve de ce fait beaucoup plus limité que celui des antédations. Nous avons ainsi postdaté trois lexèmes: *abril* (dp. 1260 vs 1182) et *agosto* (dp. 1242/1252 vs 1188/1230).

5.4. La date de dernière attestation

Lorsqu'un continuateur régulier du protoroman n'est plus employé dans un idiome roman contemporain, le DÉRom indique, lorsque c'est possible, la date de dernière attestation, la forme attestée, le cas échéant, ainsi que la source fournissant cette date. Dans notre corpus, deux lexèmes se trouvent dans ce cas: *caer*, *fêo* et *leixar*. Pour *caer*: «aport. *caer* 1278-1452/1453 [*queer*], CunhaVocabulário₂»; la date de 1452/1453 est la date la plus récente à laquelle a été trouvée une attestation de ce lexème; la note 13 précise que «cette issue régulière a été évincée par *cair* (dp. 1364 [*cqjr*])». La dernière attestation connue de *fêo* date du 15^e s. Quant à *leixar*, on le trouve jusqu'en 1552. Les dates de dernière attestation seront amenées à être corrigées, au fur et à mesure que seront trouvées des attestations plus tardives des lexèmes concernés.

D'une manière générale, toute datation peut être modifiée. Tant que le DÉRom est publié dans une version électronique, il est possible d'introduire des modifications lorsqu'elles s'avèrent nécessaires. C'est pourquoi tout article de DÉRom porte la date de mise en ligne ainsi que la date de dernière modification.

Conclusion

La démarche suivie par le DÉRom est novatrice en ce sens qu'elle a recours à un ensemble combiné de théories et de disciplines (étymologie-histoire du mot, grammaire comparée-reconstruction du protoroman, chronologie relative, géolinguistique et atlantologie,

sociolinguistique, philologie), de technologies (internet, fichiers XML) et qu'elle s'appuie sur des recherches en profondeur dans les sources bibliographiques de tous les idiomes romans, qu'il s'agisse de langues standardisées ou de dialectes. Cette démarche permet des avancées certaines tant dans le domaine panroman que dans le domaine idioroman. La comparaison entre de nombreux idiomes romans apporte un éclairage nouveau sur les étymologies d'une langue en particulier. Pour le portugais, nous avons vu que la méthode de la grammaire comparée-reconstruction permet de remonter à un étymon qui peut être différent de celui habituellement proposé, du fait de séparations post-protoromanes dues à des changements phonologiques, morphologiques ou sémantiques. Enfin, en matière de datation, la confrontation entre les différentes sources portugaises existantes et l'accès à la consultation systématique de sources galiciennes peu utilisées jusqu'à présent dans les dictionnaires étymologiques du portugais ont permis de proposer de nombreuses antédations (plus d'un lexème sur deux), des postdatations et des dates de dernière attestation. Il reste encore beaucoup à faire pour améliorer l'état de l'étymologie portugaise en matière de lexique héréditaire, beaucoup de sources à dépouiller, beaucoup d'étymons à corriger, de nouvelles datations à proposer, de nouvelles explications à chercher pour essayer de retracer l'histoire de ce lexique. Encore faut-il accepter d'inscrire ces recherches dans une perspective plus large, celle de l'étymologie panromane. Nous espérons en montrant ici ce qu'un dictionnaire comme le DÉRom peut apporter à l'étymologie portugaise, avoir pu plaider cette cause de manière convaincante.

Bibliographie

- Benarroch, Myriam (2010): *L'apport des dictionnaires de Jerónimo Cardoso (XVI^e siècle) à la datation du Dicionário Houaiss (2001)*. In: *ACILPR XXV*, 2, 623-632.
- 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. (éd.): *Quelle linguistique romane au XXI^e siècle?* Paris: L'Harmattan, 43-60.
- / 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, Frank et al. (éds.): *Congrès Mondial de Linguistique Française 2010*. Paris: Institut de Linguistique Française, publication électronique (<http://www.linguistiquefrancaise.org>), 111-123.
- / Schweickard, Wolfgang (2008): *Le Dictionnaire Étymologique Roman (DÉRom): en guise de faire-part de naissance*. In: *Lexicographica. International Annual for Lexicography* 24, 351-357.
- (2009): *Romanistique et étymologie du fonds lexical héréditaire: du REW au DÉRom (Dictionnaire Étymologique Roman)*. In: Alén Garabato, Carmen et al. (éd.): *La Romanistique dans tous ses états*, Paris: L'Harmattan, 97-110.
- (2010): *À la recherche du protoroman: objectifs et méthodes du futur Dictionnaire Étymologique Roman (DÉRom)*. In: *ACILPR XXV*, 6, 61-68.
- Buschmann Sigrid, (1965): *Beiträge zum etymologischen Wörterbuch des Galizischen*. Bonn: Romanisches Seminar der Universität Bonn.
- Castro, Ivo (2004): *A primitiva produção escrita em português*. In: *Orígenes de las lenguas romances en el Reino de León. Siglos IX-XII*. León: Centro de Estudios e Investigación San Isidoro, 69-97.

- Castro, Ivo (²2008 [¹2004]): *Introdução à história do português. Geografia da língua. Português vivo*. Lisboa: Colibri.
- Chambon, Jean-Pierre (2007): *Remarques sur la grammaire comparée-reconstruction en linguistique romane (situation, perspectives)*. In: *Mémoires de la Société de linguistique de Paris* 15, 57-72.
- (2010): *Pratique de la reconstruction 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 (éd.): *Typologie et comparatisme. Hommages offerts à Alain Lemaréchal*. Louvain / Paris / Walpole: Peeters, 61-75.
- Chauveau, Jean-Paul (2010): *Structuration des articles et rédaction du commentaire*. In: *DÉRom*. Nancy: ATILF, (<http://www.atilf.fr/DÉRom>), sous «Actualités et historique», 26-30 juillet 2010.
- Cintra, Luis F. Lindley (1971): *Nova proposta de classificação dos dialectos galego-portugueses*. In: *BF* 22, 81-116.
- (1978): *Langue parlée et traditions écrites au Moyen Âge (Péninsule Ibérique)*. In: *ACILFR XIV*, I, 463-472.
- Cunha, Antônio Geraldo da (1986-1994): *Índice do Vocabulário do Português Medieval*. Vol. 1 [A] 1986; vol. 2 [B-C] 1988; vol. 3 [D] 1994. Rio de Janeiro: Fundação Casa de Rui Barbosa.
- (1992): *Aditamento ao Índice do Vocabulário do português medieval*. In: *Confluência* 3, 23-35.
- (²2006/2007 [2002]): *Vocabulário Histórico-Cronológico do Português Medieval*, cédérom, Rio de Janeiro: Fundação Casa de Rui Barbosa / Ministério da Cultura.
- Dardel, Robert de (2007): *Une mise au point et une autocritique relatives au protoroman*. In: *RLiR* 71, 329-357.
- (2009): *La valeur ajoutée du latin global*. In: *RLiR* 73, 5-26.
- DDGM = González Seoane, Ernesto / Álvarez de la Granja, María / Boullón Agrelo, Ana Isabel (2006). *Dicionario de dicionarios do galego medieval*. Santiago de Compostela: Universidade de Santiago de Compostela [version cédérom ou site internet: <http://sli.uvigo.es/DDGM/>].
- DELP = Machado, José Pedro, (³1977 [1952]): *Dicionário etimológico da língua portuguesa*, 5 vols. Lisboa: Horizonte.
- DENF = Cunha, Antônio Geraldo da (1982): *Dicionário etimológico Nova Fronteira da língua portuguesa*. Rio de Janeiro: Nova Fronteira.
- DÉRom = Buchi, Éva / Schweickard, Wolfgang (dir.) (2008-): *Dictionnaire Étymologique Roman (DÉRom)*, site internet, Nancy: ATILF (<http://www.atilf.fr/DÉRom>).
- DRAG = Real Academia Galega (1997): *Diccionario da Real Academia Galega*. A Coruña: RAG (http://www.edu.xunta.es/diccionarios/index_rag.html).
- Emiliano, Antônio (1999): *O mais antigo documento latino-português (882 a.D.)*. Edição e estudo grafêmico. In: *Verba* 26, 7-42.
- (2003): *Sobre a questão d'«os mais antigos textos escritos em português»*. In: Castro, Ivo / Duarte, Inês (edd.): *Razões e Emoção. Miscelânea de estudos em homenagem a Maria Helena Mira Mateus*. Lisboa: Imprensa Nacional – Casa da Moeda, 261-278.
- Hall, Robert A. Jr (1950): *The reconstruction of proto-romance*. In: *Language* 26, 6-27.
- Herculano, Alexandre (1867): *Portugaliae Monumenta Historica a Saeculo Octavo post Christum usque ad Quintum Decimum. Diplomata et Chartae*. Lisboa: Academia Real das Ciências.
- Houaiss, Antônio / Villar, Mauro de Salles (²2007 [¹2001]): *Dicionário eletrônico Houaiss da língua portuguesa*, cédérom (2.0a), Rio de Janeiro: Objetiva.
- LB = *Livre Bleu* (version du 18/19.11.2011). In: *DÉRom (Dictionnaire Étymologique Roman)*. Nancy: ATILF, site internet (<http://www.atilf.fr/DÉRom>), accès réservé (ouvert sur demande motivée).
- Lisboa, Eduardo de (1937): *O Dicionário do Sr. Nascentes e o REW. Rectificações*. Rio de Janeiro: Pimenta de Mello.
- Lorenzo, Ramón (1968): *Sobre cronologia do vocabulário galego-português (Anotações ao Dicionário etimológico de José Pedro Machado)*. Vigo: Galaxia.

- (1995): *La koiné galega (Galegische Koine)*. In: *LRL* II, 2 (n° 160), 649-679.
- (2007): *Os notarios e a lingua nos comezos da escrita documental en galego*. In: Boullón Agrelo, Ana Isabel (ed.): *Na nosa lingoage galega. A emerxencia do galego como lingua escrita na Idade Media*. Santiago de Compostela: CCG / ILG, 313-372.
- Maia, Clarinda de Azevedo (1986): *História do Galego-Português. Estado lingüístico da Galiza e do Noroeste de Portugal desde o século XIII ao século XVI (com referência à situação do galego moderno)*. Lisboa: Fundação Calouste Gulbenkian / Junta Nacional de Investigação Científica e Tecnológica.
- Martinez Salazar, Andrés (1911): *Documentos gallegos de los siglos XIII a XVI*. A Coruña: Casa da Misericórdia.
- Martins, Ana Maria (1999): *Os mais antigos textos escritos em português: Documentos de 1175 a 1252*. In: Hub Faria, Isabel (ed.): *Lindley Cintra. Homenagem ao Homem, ao Mestre e ao Cidadão*. Lisboa: Cosmos / Faculdade de Letras da Universidade de Lisboa, 491-534.
- (2007): *O primeiro século do português escrito*. In: Boullón Agrelo, Ana Isabel (ed.): *Na nosa lingoage galega. A emerxencia do galego como lingua escrita na Idade Media*. Santiago de Compostela: CCG / ILG, 161-184.
- Monjour, Alf (2004): *El diccionario Houaiss y la etimología portuguesa*. In: Boullón Agrelo, Ana Isabel (ed.): *Novi te ex nomine. Estudos filológicos ofrecidos ao Prof. Dieter Kremer*. A Coruña: Fundación Pedro Barrié de la Maza, 145-155.
- Mattos e Silva, Rosa Virgínia (2008): *O português arcaico: uma aproximação*. 2 vols. I: Léxico e morfologia; II: Sintaxe e fonologia, Lisboa: Imprensa Nacional-Casa da Moeda.
- Pereira, António (2002). *Alguns d'«os mais antigos textos escritos em português» «Notícia de fiadores» (1175): estudo antroponímico*. In: *Actas do Encontro Comemorativo dos 25 anos do Centro de Linguística da Universidade do Porto* 1, 69-82.
- Piel, Joseph M. (1932): *Notas à margem do «Romanisches Etymologisches Wörterbuch»*. In: *Biblos* 8, 379-392.
- REW = Meyer-Lübke, Wilhelm (³1930-1935 [1911-1920]): *Romanisches Etymologisches Wörterbuch*. Heidelberg: Winter.
- Sacks, Norman P. (1941): *The Latinity of dated documents in the Portuguese territory*. Philadelphia: University of Pennsylvania.
- Souto Cabo, José António (2004): *Novas perspectivas sobre a génese da scripta romance na área galego-portuguesa. Textos e contextos*. In: *Aemilianense* I, 569-599.
- Teyssier, Paul (1982): *História da Língua portuguesa*. Lisboa: Sá da Costa (¹1980, Paris, PUF, pour l'édition originale).
- (1995): *La koiné portugaise (Portugiesische Koine)*. In: *LRL* II, 2 (n° 161), 679-692.
- Vasconcelos, José Leite de (1928): *Opúsculos* I, Coimbra, Imprensa da Universidade.
- (1987 [1901]): *Esquisse d'une dialectologie portugaise*. Lisboa, CNIC.