

## Structuration étymologique du lexique galloroman avant le 12<sup>e</sup> siècle

### 1. Introduction

L'aperçu suivant réunit quelques résultats d'un travail en cours qui porte sur le lexique galloroman de la première époque<sup>1</sup>. Plus précisément, nous étudions les éléments vernaculaires qui sont présents dans les documents latins originaux rédigés en France entre le 8<sup>e</sup> et le début du 12<sup>e</sup> siècle. Nous souhaiterions aborder ici un aspect plus spécifique de notre recherche, à savoir la question de la régionalité du lexique galloroman et la structure étymologique qui lui est sous-jacente. Il s'agit de pistes interprétatives provisoires reposant sur une partie de la nomenclature que nous sommes en train de traiter.

Il est un fait que personne n'a pu étudier jusqu'ici sur une base empirique la diversification régionale du lexique galloroman avant l'apparition des premiers textes vernaculaires pleins vers 1100. La tradition lexicographique du français et de l'occitan, bien qu'exemplaire, connaît une lacune patente pour la documentation antérieure au XII<sup>e</sup> siècle ; en d'autres termes, elle débute presque 500 ans après la formation des langues romanes. Or, pour pouvoir dépasser cette lacune, la seule possibilité est de travailler sur les textes rédigés en latin. Ceux-ci contiennent un lexique vernaculaire pré-textuel bien présent mais qui n'a jamais été étudié de manière très systématique.

Nous avons pu nous rendre compte de la pertinence des éléments vernaculaires en contexte latin à l'occasion de notre travail de thèse sur l'émergence de l'occitan auvergnat (Carles 2011). Il s'agit de matériaux fiables, bien que fragmentaires, donnant un accès inédit à la réalité linguistique romane de cette époque. Dans le cadre de notre projet d'habilitation, nous étudions depuis maintenant quatre ans les lexèmes gallo-romans en contexte latin.

Cette entreprise a été rendue possible grâce à la mise en ligne en 2010 du corpus des *Chartes originales antérieures à 1120 conservées en France* réunies et transcrites par l'ARTEM. La lecture des quelques 5000 documents nous a permis de relever un nombre important de lexèmes français, occitans et dans une moindre mesure, francoprovençaux.

---

<sup>1</sup> Nous avons déjà eu l'occasion, lors du dernier congrès de la Société de Linguistique Romane à Valence, de présenter les objectifs de notre projet qui en était alors à ses débuts (Carles 2013).

Nous avons ensuite, sur la base de notre premier relevé, identifié des attestations complémentaires grâce aux fonctions de recherches informatiques de la base de l'ARTEM. À partir d'un relevé d'un millier de formes d'un intérêt vernaculaire, nous avons retenu environ 600 lexèmes qui font l'objet d'un traitement approfondi dans un cadre lexicographique. Ce 'Trésor galloroman des origines' constitue la partie centrale de notre travail.

## 2. Questions de méthodologie

Nous avons étudié à ce jour (juillet 2013) 189 lexèmes occitans, français ou franco-provençaux qui se placent dans 160 articles rédigés. Nous avons traité les lettres A M, N, O et P qui représentent environ un tiers de notre nomenclature finale.

La plupart de ces lexèmes sont des substantifs, parfois des adjectifs, rarement des verbes. Nous avons également intégré les témoignages précieux livrés par l'onomastique en exploitant les toponymes délexicaux à article (du type *la Forge*). La présence de l'article défini garantit en effet que ces noms propres sont de formation romane. De même, nous avons retenu certains surnoms en voie d'onymisation : par exemple *Vivianus li Mechs* relevé dans le Cher à la fin du 10<sup>e</sup> siècle.

Une fois identifié un mot à caractère vernaculaire dans un document original, avant tout par des marques grapho-phonétiques, nous avons élargi notre recherche dans des actes contemporains copiés. Nous avons interrogé notamment la base de données des *Chartae Galliae* qui réunit 36000 chartes transmises le plus souvent en copie et antérieures à 1300. L'utilisation de cette masse documentaire a notamment permis de réduire le danger de contingence de nos données originales.

Notre objectif a été ensuite de cerner l'axe diachronique des lexèmes en question et, parallèlement, de déterminer leur distribution géolinguistique dans la Galloromania. Nos choix méthodologiques ressortent de la lecture de l'article suivant, AFFIZARE.

apoit. lat. AFFIZARE v.tr.

(apoit. lat. *affidare* attesté entre 1082 et 1095, *affizare* 1087/1100)

♦ t. juridique, v.tr. *affizare* « s'engager à réaliser ou à faire réaliser quelque chose, fr. garantir ».

1087/1100 « Quod si famulus monachorum annonam conduxerit alicujus ad molendinum rogatus vel remuneratus, ipsum et annonam et asinos, quamdiu in ejus conductu fuerint, similiter per omnia *affizamus*. (...) Si autem aliquis presumpserit ipsos molendinos infringere, vel que infra fuerint invadere, vel famulis monachorum et que in conductum habuerint ad molendinum venientes aliquid nocere, faciemus illis justiciam competentem, et hoc *affizamus* et promittimus a nobis et a nostris ita in perpetuum propter amorem Dei custodiri, ut et ipsi monachi pro nobis et pro amicis nostris vivis et defunctis generales faciant orationes » (promesse de protection de moulins faite par un seigneur au profit de l'abbaye de Saint-Cyprien de Poitiers, AD Vienne, C 12 dossier 3 n°15, ARTEM n° 1321 [or.]).

*Données supplémentaires* : le corpus de l'ARTEM donne plusieurs attestations de forme pleinement latinisée (antérieures à Niermeyer et DC cf. *infra*) : *affidare* 1082 (n° 3434, Maine-et-Loire), 1083/86 (n° 3258, Maine-et-Loire), 1095 (n° 3397 et 3398 *id.*). Les *Chartae Galliae* enregistrent une cinquantaine d'attestations de la forme latinisée *affidare* en domaine d'oil. — Niermeyer s.v. *affidare* atteste le verbe dans le sens 1. « promettre fidélité » depuis le 12<sup>e</sup> s. sans localisation ; DC s.v. *affidare* donne comme première attestation une charte marseillaise de 1193.

Il s'agit du continuateur d'un verbe préfixé probablement d'époque latine \*AFFIDARE formé sur \*FIDARE.

Le type aocc. est attesté par FEW 3, 500b, \*FIDARE : « apr. *affizar* « assurer, garantir » (Lv ; Jaufre), abéarn. *affidar* ». Il est toutefois impossible d'identifier la source de Levy puisque le verbe en question n'est mentionné que dans LvP (absent dans Lv). Rn 3, 290 atteste le verbe en contexte documentaire en 1068 (*affidar*), puis dans le *Roman de Fierabras*, chez Arnaut de Marueil et Giraut de Borneil. Absent dans Brunel et DAOA. FEW *ib.* atteste le type en français entre le 12<sup>e</sup> et le 16<sup>e</sup> s. : « afr. mfr. *afier* [« assurer, promettre, donner en gage »] (12.–16. jh., Gdf ; TL ; Arch 141, 205 ; Guill Pal ; Gace ; Proth ; BenSmaure) ».

La localisation des attestations dialectales oïliques dans des sens proches ne révèle pas de diffusion diatopique spécifique. Pour le poitevin, le FEW n'indique pas d'attestations anciennes mais seulement des formes modernes de la région concernée (Gâtine, Deux-Sèvres et Saintonge) qui appartiennent au type oïlique *affier* (de même sens). Cf. FEW 4, 500b ; 502 n. 20.

La forme de 1087/1100 est mixte (élément vernaculaire accompagné d'une régularisation morphologique latine). Cette occurrence illustre un passage à la sifflante sonore de /d/ intervocalique primaire noté <z> et prouve du même coup l'usage poitevin du type occitan *affizar*, jusqu'ici non répertorié. Le phonétisme occitan reste donc présent dans cette région au moins jusqu'au 12<sup>e</sup> siècle avant d'être définitivement supplanté par la formation parallèle oïlique *afier* (avec amuïssement de la dentale intervocalique).

Comme entrée figurent le lemme des formes relevées dans le corpus de l'ARTEM (ici, apoit. lat. AFFIZARE v.tr.). Pour déterminer l'étiquette de langue, nous prenons en considération les acteurs des chartes, les toponymes mentionnés dans les documents et le lieu de conservation des actes.

A la ligne suivante, se trouve une synthèse des attestations en original tirées du corpus de l'ARTEM (ici, apoit. lat. *affidare* attesté entre 1082 et 1095, *affizare* 1087/1100). Vient ensuite la partie consacrée à la description du lexème : une éventuelle marque diasystématique, la catégorisation grammaticale, le lemme puis sa définition dans les contextes relevés.

Nous indiquons ensuite les attestations : il s'agit ici d'un texte daté entre 1087 et 1100 conservé en original, concernant des biens localisés à Cragon, qui est un ancien domaine de l'abbaye Saint-Cyprien de Poitiers situé dans la commune de Saint-Jean-de-Sauve. Un seigneur garantit ici la protection des moulins de Cragon appartenant aux moines de Saint Cyprien.

Dans la rubrique suivante, nous mentionnons des attestations complémentaires de l'ARTEM, en l'occurrence ici les attestations de forme pleinement latinisée (mais non attestée en latin antique) : on en relève deux en Anjou. En ce lieu, on trouve également les données pertinentes relevées dans les *Chartae Galliae*, les données de Niermeyer et celles de Du Cange.

Suivent l'étymon, puis le commentaire géolinguistique et diachronique et, enfin, les observations grapho-phonétiques. Le traitement de ce mot met donc au jour une attestation poitevine médiévale, qui présente un traitement phonétique occitan.

### 3. Les catégories des lexèmes vernaculaires en contexte latin

Avant de pouvoir appréhender la régionalité de notre lexique, il est indispensable de le catégoriser d'après ses époques de formation.

Dans l'étude de ces mots nous prenons en considération deux paramètres fondamentaux : (1) leur époque de formation et (2) leur tradition textuelle. Si on ne saurait nier que l'époque de formation est importante pour l'histoire du lexique, on admettra qu'il s'agit là d'un paramètre créé par le linguiste à ses propres fins. En effet, la distinction entre un mot d'origine gauloise, une formation latine ou un dérivé d'époque romane n'est pas pertinente au moment de la mise à l'écrit de ces mots au 10<sup>e</sup> siècle. Elle n'intéressait pas, et pour cause, le scribe du Haut Moyen Âge. La tradition textuelle d'un mot est en revanche un critère déterminant pour les auteurs des actes médiévaux. Si un lexème connaît une bonne documentation à l'écrit dans l'Antiquité, son emploi dans les chartes repose sur une tradition déjà ancienne. Dans le cas contraire, les scribes doivent transposer la forme en question à partir de l'oral.

Nous distinguons par conséquent deux grands ensembles : les mots vernaculaires dont les étymons sont bien attestés en latin et ceux qui ne le sont pas. Dans le cas des mots bien attestés, les scribes médiévaux ont le choix entre une forme latine de tradition ancienne et sa forme vernaculaire correspondante ; en l'absence d'une telle tradition, ils choisissent entre une transposition écrite vernaculaire de la forme orale et une conversion secondaire de celle-ci en latin (relatinisation)<sup>2</sup>.

#### 3.1. *Lexèmes à tradition écrite antique*

Nous évoquons tout d'abord la catégorie des mots dont l'étymon est bien attesté en latin antique et qui n'a pas connu de dérivation ou de changement sémantique postérieur. Ces mots de notre corpus se démarquent de l'entourage latin des actes seulement par le fait d'entériner des changements phonétiques romans. Quelques exemples :

afr. champ. *marrin* s.m. « bois de construction » 1101 < lat. MATERIAMEN s.n.

<sup>2</sup> Dans les deux cas, ils peuvent encore opter pour une voie intermédiaire en créant une forme 'mixte' comme *affizare* intégrant à la fois des éléments latins et vernaculaires. Notre travail de thèse a montré que ce cas de figure est caractéristique de l'émergence du vernaculaire à l'écrit.

aocc. *paissel* s.m. « pieu en bois que l'on enfonce au pied d'un cep de vigne pour le soutenir » 1017 < lat. *PAXILLUS* s.m.

Ce cas de figure est relativement fréquent avec presque 30 % des lexèmes étudiés jusqu'ici. Les attestations se placent surtout après l'an 1000 et apparaissent ainsi à une époque avancée de l'émergence du vernaculaire à l'écrit. Elles permettent néanmoins d'antidater souvent considérablement les données du FEW. Par exemple, si le type oïlique continuateur de *MATERIAMEN* est bien attesté depuis le 12<sup>e</sup> siècle, le traitement phonétique champenois *marrin* n'est enregistré dans le FEW qu'à partir du milieu 15<sup>e</sup> siècle (*ib.*, 486b).

On notera que les toponymes à article sont particulièrement fréquents dans cet ensemble de mots.

Malgré son apparente simplicité, cette catégorie demandera une analyse détaillée, une fois l'étude de la nomenclature achevée.

### 3.2. *Mots d'origine gauloise, mots préromans et onomatopées*

Pour toutes les autres catégories, l'étymon n'est pas attesté dans l'Antiquité, ou seulement ponctuellement et tardivement. On distinguera parmi eux dans la mesure du possible les formations vraisemblablement d'époque latine et celles d'époque romane. Comme point de bascule théorique entre latin et roman, nous avons retenu en tant qu'hypothèse de travail la date de *ca* 700 proposée par Chambon 2005 sur la base du critère de l'introduction de l'article défini. Cette date ne met pas en cause l'idée du continuum linguistique, mais permet de distinguer deux époques de formation et également de tradition écrite dont l'opposition peut s'avérer utile. Nous reviendrons plus loin sur cette distinction (cf. *infra* 3.4).

L'ancienneté de la formation est assurée pour les mots d'origine gauloise puisqu'ils ont été intégrés au latin parlé avant le 5<sup>e</sup> siècle. Nous relevons jusqu'ici sept celtismes ou mots préromans, tous non dérivés, auxquels s'ajoute une onomatopée et un mot d'origine inconnue. On peut noter que les mots de cette catégorie connaissent souvent des attestations très anciennes dans notre corpus ; par exemple :

afr. lat. *anzinga* s.f. « mesure agraire pour les terres labourables, les vignes et les bois » 854 < lat. \**ANDEKINGA* < gaul. \**ande-cingo*

### 3.3. *Mots d'origine germanique*

#### 3.3.1. *Emprunts non dérivés*

Parmi les germanismes, il convient de distinguer les formes avec et sans dérivation. Les formes sans dérivation correspondent à des emprunts directs dont l'époque se place pour la plupart avant 700 ; plusieurs de ces emprunts sont d'ailleurs ponctuellement attestés en latin avant cette date. C'est le cas par exemple de :

afr. *plege* s.m. « personne qui se donne comme caution » 1051 < lat. tard. *PLEBIUM* 595 < germ. \**plegan*

Nous en relevons à ce jour sept dans notre corpus, dont les attestations se placent exclusivement au 11<sup>e</sup> siècle.

### 3.3.2. Dérivés formés à partir d'emprunts germaniques

En revanche, les dérivés formés à partir de tels emprunts germaniques appartiennent plus probablement à l'époque romane (8 lexèmes) ; par exemple :

afr. *aloter* s.m. « possesseur d'un fonds de terre possédé en pleine propriété, exempt de droits féodaux » 1037/45 < lat. tard. *ALODE* 507/11 < germ. \**alôd*

### 3.4. Mots dérivés à base latine sans tradition écrite antique

L'ensemble le plus important enfin est constitué de mots dérivés à base latine dont l'étymon n'est pas attesté. Il peut s'agir de formations latines ou bien de formations romanes. Les deux types confondus correspondent à presque la moitié des 189 lexèmes traités à ce jour.

La distinction entre les formations latines et romanes est naturellement très délicate. Nos critères opératoires sont, dans une combinaison variable,

- (i) la chronologie relative des changements phonétiques – critère surtout porteur en domaine d'oïl –
- (ii) la vitalité chronologique des suffixes,
- (iii) la chronologie des attestations médiévales,
- (iv) la diffusion large ou restreinte du lexème dans Galloromania et la Romania, enfin
- (v) la vitalité et la productivité dérivationnelle et sémantique des lexèmes en question à l'époque médiévale.

Sur la base de ces critères, nous avons attribué, à ce stade de réflexion, une cinquantaine de lexèmes à l'époque latine et une quarantaine à l'époque romane.

Parmi les formations supposées d'époque latine, on peut citer pour l'occitan et le français les lexèmes suivants :

aocc. *parran* s.m « jardin près d'une maison entourée d'une clôture » 1078/1108 < lat. \**PARRAGINE*

aocc. *perair* s.m. « arbre fruitier de taille moyenne, donnant des poires » mil. 11<sup>e</sup> s. < lat. \**PIRARIUS*

afr. *molnar* s.m. « emplacement du moulin » [954/55], 1060/70 < \**MOLINARE*

afr. *pitit* s.m. « homme qui est d'une taille inférieure à la moyenne » extrait de NP 1115 < \**PETTITUS*

afr. lat. *mallia* s.f. « monnaie de billon au-dessous du denier » 1067/1110 < \**MEDALIA*

Parmi les formations supposées d'époque romane, mentionnons,

aocc. *meissonenc* adj. « au moment de la moisson » 1047/60 < aocc. *meisso* < *MESSIO*

aocc. *preveiril* adj. ou s.m. « (fief) attribué au curé d'une paroisse » 1100/25 < aocc. *preveire* < PRESBYTER

aocc. lat. *achaptamentum* s.m. « acquisition de la part seigneuriale dans un contrat emphytéotique » 1031/60 < aocc. *achaptar* < \*ACCAPTARE

afr. *paisant* s.m. « personne de la campagne qui vit de la culture du sol et de l'élevage des animaux » 1082/89 < afr. *pais* + *-enc* < PAGENSIS

afr. lat. *mariagium* s.m. « bien qu'apporte une femme en se mariant » 1061 < afr. *marier* < MARITARE

De manière générale on observe pour les dérivés dont l'étymon n'est pas attesté en latin une certaine disparité quantitative entre le nord et le sud de la Galloromania. Les mots occitans sont plus fréquents, avec 60 %, que les mots français avec 40 %. Si le nord a donc une certaine tendance à freiner d'avantage l'élément vernaculaire à l'écrit, les deux ensembles s'inscrivent néanmoins dans des logiques comparables ; il s'agit d'une différence de degré et non pas de nature.

Ce qui est plus flagrant, c'est la tendance du nord à relatiniser plus volontiers l'apparition des éléments vernaculaires : environ deux tiers des formes françaises sont partiellement latinisées contre seulement un quart des formes occitanes.

Il est permis de se demander si cette disparité est imputable, entre autres, à une question de distance linguistique. Indéniablement la distance entre le latin et l'occitan est moins prononcée que celle entre le latin et le français, moins conservateur. Il est également patent que les scribes de cette période charnière sont peu enclins à rompre l'harmonie textuelle latine. C'est ce qui explique le nombre important de formes mixtes, partiellement latinisées. En tout état de cause, l'hypothèse d'incompétence des scribes du sud doit être écartée.

#### 4. La régionalité lexicale

La structure étymologique et chronologique de notre nomenclature permet d'aborder la question de la régionalité lexicale sous un angle diachronique. Notre corpus contient des termes pangalloromans et des lexèmes qui appartiennent seulement au domaine d'oc ou au domaine d'oïl. Par ailleurs, une grande partie des mots traités connaît des restrictions géolinguistiques à l'intérieur de la Galloromania qui ne saurait se réduire à l'opposition entre oc et oïl.

Nous présenterons donc ces différentes catégories plus en détail pour les 189 lexèmes déjà traités :

##### 4.1. Les mots pangalloromans

Le premier constat assez surprenant est que la moitié des lexèmes connaît une distribution pangalloromane. Ceux-ci ne présentent aucune régionalité reconnaissable, ni dans notre corpus ni d'après les données du FEW. Ce nombre élevé montre la

grande cohérence lexicale qui existe entre les langues galloromanes au moins d'après le témoignage des mots vernaculaires en contexte latin à notre époque.

4.1.1. En outre, les deux tiers de ces mots sont des mots hérités du latin, presque exclusivement non construits et sans dérivation ultérieure. Par exemple,

aocc. *obra* « journal de terre » 1017/21 < lat. *OPERA* s.f.

4.1.2. Parmi le tiers restant, les dérivés à diffusion pangalloromane, la grande majorité semble être de formation antérieure à 700, à l'exemple de *pereir*, déjà mentionné :

aocc. *pereir* s.m. « arbre fruitier donnant des poires » mil. 11<sup>e</sup> s. < lat. \**PIRARIUS*

4.1.3. On relève toutefois dans cette catégorie une série de formations d'époque romane, qui sont parallèles en français et en occitan, par des évolutions convergentes. C'est le cas d'aocc. *mezalhada* s.f. « mesure de terre (*ici* : s'appliquant à la vigne) », formé sur *mezalha* s.f., parallèle à afr. *mailhee* s.f., de même sens, qui est une formation bien plus tardive (1375 et 1471) sur afr. *maille* s.f. (cf. FEW 6<sup>1</sup>, 573a, *MEDIALIS*).

#### 4.2. Les mots d'oc

Une petite proportion de mots – un peu moins de 10% – n'appartient qu'au domaine d'oc et est inconnue du domaine d'oïl. Pour ces lexèmes, il n'est pas possible de déceler une régionalité en occitan. Précisons qu'il ne s'agit pas de la distribution géolinguistique observée dans notre corpus mais celle qui se dessine dans les articles du FEW.

Le premier constat à faire, est que ces mots représentent exclusivement des dérivés. Par ailleurs, presque la moitié est de formation postérieure à 700, alors que ce n'était le cas que pour un dixième seulement des mots pangalloromans. C'est le cas par exemple, d'aocc. lat. *achaptamentum* s.m. « acquisition de la part seigneuriale dans un contrat emphytéotique » (1031/60). Le substantif est formé à partir du verbe *achaptar* (< \**ACCAPTARE*) et n'est attesté en domaine d'oc qu'aux 11<sup>e</sup> et 12<sup>e</sup> s.

#### 4.3. Les mots d'oïl

Le nombre de lexèmes inconnus de l'occitan est plus important, avec près de 15%. La proportion de mots construits est de nouveau écrasante et les formations d'époque romane concernent encore presque la moitié des formes exclusivement oïliques. Par exemple,

afr. lat. *astalagium* s.n. « redevance due sur les marchandises étalées » 1076, < afr. *estal* s.m. < anfrq. \**stal*

afr. lat. *mariagium* « bien qu'apporte une femme en se mariant » 1061 < afr. *marier* < *MARITARE*



#### 4.4. Les lexèmes à diffusion régionale

Pour les lexèmes restants, la distribution géolinguistique n'est ni pangalloromane, ni restreinte aux domaines d'oc OU d'oïl. Ils représentent près de 30% de la nomenclature traitée [54].

On distingue trois grands groupes :

4.4.1. Un premier groupe correspond à des lexèmes occitans dont l'aire de diffusion dépasse le territoire occitan vers le nord en incluant le sud du territoire d'oïl et/ou le francoprovençal. C'est le cas de

apoit. lat. *mulnarius* 954/55, afrpr. lat. *mulnarius* 957, aang. *molnar* 1060/70 < \*MOLINARE

Notre corpus met ici en évidence que *molnar* « emplacement du moulin » est présent en Anjou et en domaine francoprovençal à date ancienne ; l'attestation poitevine (954/55) indique par ailleurs que la région a, sans surprise, partagé un temps le type occitan. Par la suite, ce lexème connaît une régression vers le sud (cf. FEW 6<sup>3</sup>, 40a, MOLINUM).

Dans trois cas ensuite, la répartition est plutôt orientale (en incluant le francoprovençal : *masura*, *pascherium*, *paxeria*), dans cinq cas plutôt occidentale, avec une concentration toute particulière dans le Poitou (*arbergamentum*, *affizare*, *albergaria*, *mec*, *oblia*). La plupart de ces mots sont de formation antérieure à 700. Il s'agit donc de solidarités géolinguistiques anciennes et non pas d'évolutions parallèles ou convergentes.

4.4.2. Parallèlement à ce premier groupe, un deuxième réunit des lexèmes oïliques dont l'aire de diffusion s'élargit vers le domaine d'oc. Nous relevons quatre termes qui sont partagés par l'occitan septentrional (*aissimentum*, *arzilerius*, *masel*, *nauda*), un à l'occitan oriental (*polveradge*) et un à l'occitan septentrional et oriental (*moiada*). La plupart de ces formations sont également d'époque latine.

4.4.3. Un dernier grand ensemble montre enfin des solidarités géolinguistiques plus complexes, dans lesquelles l'opposition entre oc et oïl n'est pas pertinente. Sont concernés 37 mots répartis en 28 types lexicaux que le tableau suivant tente de catégoriser de manière succincte :

– une seule région :	Auvergne :	<i><u>apradat</u></i> , <i>pedrucia</i> , <i>moleta</i> , <i>parceria</i> , <i>masale</i> , <i>monsteriole</i>
	Languedoc :	<i><u>amasadus</u></i> , <i>mercadal</i>
	Centre :	<i><u>menbrerola</u></i> , <i>palagium</i> (sém.)

– deux régions occitanes voisines :	Auvergne / Languedoc :	<i><u>abseddat</u>, nava</i>
	Limousin / Languedoc :	<i>preveiril</i>
	Languedoc or. / Provence :	<i><u>affrontar</u>, <u>afrotaciones</u>, <u>meissonenc</u></i>
– un territoire d’oc ou d’oïl plus large :	Languedoc / Provence :	<i><u>melgoiresense</u>, <u>mergolies</u>, <u>porcada</u></i>
	Auvergne / Limousin / Languedoc sept. / Provence :	<i>acaptar</i>
	Poitou / Centre / Ouest :	<i>aguot</i>
– une zone intermédiaire N/S :	Centre / Normandie :	<i>marra</i>
	occ. sept. / frpr. :	<i>apendaria</i>
	occ sept. / Poitou / Centre / Ouest :	<i>apradar, pledura</i>
– une grande zone N/S :	zone occidentale :	<i>masura</i>
	zone orientale (prov. / bourg. / frpr.) :	<i><u>alberga</u>, <u>peciola</u></i>

Certains sont des diatopismes restreints à une seule région (par exemple l’Auvergne), d’autres concernent deux régions limitrophes (par exemple le Languedoc oriental et la Provence). Il s’agit parfois de franges plus importantes qui se dessinent dans le FEW, faisant écho à nos données plus anciennes.

La distribution étymologique de ces 28 types lexicaux est très caractéristique : seulement deux ont un étymon bien attesté depuis l’Antiquité. Toutes les autres formes proviennent soit de l’oralité latine d’avant 700 soit de l’oralité romane.

Aussi on remarque que la quasi-totalité des dérivés romans correspond à des diatopismes limités à une région voire à deux régions voisines (ces formes sont soulignées dans le tableau).

A l’inverse, l’extension géolinguistique est plus large pour les formations antérieures à 700. C’est le cas de :

aocc. *peciola* s.f. « portion de terre d'un seul tenant et appartenant à un individu » 1008/11<sup>e</sup> s., afr. bourg. lat. *petiola* 907–1023, afrpr. lat. *petiola* 951/1031, afrpr. lat. *peciola* 942/54–1037 < lat. \*PETTIOLA < lat. \*PETTIA < gaul. \*pettia

L'aire de ce lexème englobe les domaines bourguignon, francoprovençal, provençal jusqu'en Italo-romania (cf. FEW 8, 332sqq., \*pettia).

Il apparaît assez clairement que la répartition géolinguistique des mots de notre corpus dont l'étymon n'est pas attesté en latin avant 700 est assez protéiforme, tandis que celle des mots à l'étymon bien attesté s'inscrit dans des logiques plus simples.

Une fois notre nomenclature achevée, il sera notamment intéressant d'étudier de près la position géo-historique du poitevin et du francoprovençal, car ils connaissent une présence significative dans notre corpus.

De manière plus générale, il est d'ores et déjà reconnaissable que le *Trésor gallo-roman des origines* permettra de préciser le développement et le degré de la régionalité lexicale galloromane des 8<sup>e</sup>-12<sup>e</sup> siècles.

## 5. Conclusions

L'avancement actuel de nos recherches permet, en synthèse, les observations suivantes :

D'une part, les mots vernaculaires relevés dans les documents latins antérieurs à 1120 montrent que la cohérence lexicale de l'espace galloroman est indéniable, puisqu'elle concerne 50% des cas. Néanmoins, l'opposition entre domaine d'oc et domaine d'oïl est également bien reconnaissable (20%). Enfin, nos attestations anciennes conjointes aux données du FEW permettent de mieux saisir la régionalité lexicale à date ancienne, notamment celle qui ne se restreint pas à la binarité oc/oïl et qui concerne presque un tiers des lexèmes.

Nous avons pu constater par ailleurs que la complexité de la répartition géolinguistique est nettement plus forte parmi les formes de l'oralité latine que parmi les mots avec une tradition textuelle déjà antique. Cette complexité augmente encore pour les mots de formation romane.

Nous avons ainsi l'espoir que notre étude ouvrira la voie à une vue d'ensemble structurante sur un usage linguistique à ce jour inconnu. Il nous semble notamment important de mettre fin à l'idée trop largement répandue selon laquelle aux 8<sup>e</sup>-10<sup>e</sup> siècles les langues galloromanes seraient des 'protolangues'. L'analyse de nos matériaux lexicaux démontrera sur une nouvelle base empirique que les variétés galloromanes de cette époque sont pleinement formées, régionalisées et diversifiées quant à leurs modes d'expression.

## Bibliographie

- Adams, James Noel, 2007. *The Regional Diversification of Latin 200 BC – AD 600*, Cambridge, Cambridge University Press.
- ARTEM, 2010. *Actes originaux antérieurs à 1120 conservés en France*, publication en ligne (<medievistique.univ-nancy2.fr/contentId%3D7911>).
- Carles, Hélène, 2011. *L'émergence de l'occitan pré-textuel. Analyse linguistique d'un corpus auvergnat (IX<sup>e</sup>-XI<sup>e</sup> s.)*, Strasbourg.
- Carles, Hélène, 2013. « Pour un *Trésor Galloroman des Origines*: les lexèmes vernaculaires et les toponymes délexicaux dans les plus anciennes chartes originales latines », in: Cesáreo Calvo/Emili Casanova (éd.), *Actes del XXVI<sup>e</sup> Congrés de lingüística i filologia romàniques, València 2010. Secció 4, Lexicologia et Fraseologia*, Berlin, De Gruyter, 39-50.
- Chambon, Jean-Pierre, 2005. « Toponymie et grammaire historique: les noms de lieux issus de *cappella* et *forestis* et la diffusion spatiale de l'article défini dans la Galloromania », in: James Raoul, Danièle/Soutet, Olivier (éd.), *Par les mots et les textes. Mélanges (...) offerts à Claude Thomasset*, Paris, PUPS, 143-155.
- Chambon, Jean-Pierre, 2007. « Remarques sur la grammaire comparée-reconstruction en linguistique romane (situation, perspectives) », in: Lemaréchal, Alain (ed.), *Tradition et rupture dans les grammaires comparées de différentes familles de langues* (Mémoires de la Société de Linguistique de Paris, 15), 57-72, Louvain.
- Chambon, Jean-Pierre/Carles, Hélène, 2005. « Un censier de Plauzat (Basse Auvergne) partiellement rédigé en langue d'oc », *Lengas* 58, 7-42.
- Koch, Peter, 1993. « Pour une typologie conceptionnelle et médiale des plus anciens documents/monuments des langues romanes », in: Selig/Frank/Hartmann 1993, 39-81.
- Lüdtke, Helmut, 2009 [2005]. *Der Ursprung der romanischen Sprachen: eine Geschichte der sprachlichen Kommunikation*, Kiel, Westensee-Verlag.
- Menéndez-Pidal, Ramón/Lapesa, Rafaël/Seco, Manuel, 2003. *Léxico hispanico primitivo (siglos VIII al XII)*, Madrid, Fundación Ramón Menéndez Pidal.
- Pitz, Martina, 2008. « L'amuïssement des voyelles finales en proto-français. Indices chronologiques fournis par la toponymie d'époque mérovingienne », in: Dahmen, Wolfgang et al. (ed.), 2008. *Zur Bedeutung der Namenkunde für die Romanistik. Romanistisches Kolloquium XXII*, Tübingen, Narr, 229-249.
- Selig, Maria/Frank, Barbara/Hartmann, Jörg (éd.), 1993. *Le passage à l'écrit des langues romanes*, Tübingen, Narr.
- Tock, Benoît-Michel et al. (éd.), 2001. *La diplomatie française du Haut Moyen Age. Inventaire des chartes originales antérieures à 1121 conservées en France*, Turnhout, Brepols.
- Van Acker, Marieke/Van Deyck, Rika/Van Uytfanghe, Marc (ed.), 2008. *Latin écrit – Roman oral? De la dichotomisation à la continuité*, Turnhout, Brepols.
- Vitali, David, 2007. *Mit dem Latein am Ende? Volkssprachlicher Einfluss in lateinischen Chartularen aus der Westschweiz*, Berne, Lang.
- Wüest, Jakob, 1979. *La Dialectalisation de la Gallo-Romania*, Berne, Francke.