Le Thesaurus Occitan : entre atlas et dictionnaire

Patrick Sauzet, Guylaine Brun-Trigaud

Le Thesaurus Occitan : entre atlas et dictionnaires

Patrick SAUZET
Université de Toulouse & CNRS « Cognition, Langues, Langage, Ergonomie » [1]
Guylaine BRUN-TRIGAUD
C.N.R.S. « Bases, Corpus et Langage » (Nice)[2]

Résumé : Les études occitanes souffrent d'un manque de chercheurs. Les causes en sont multiples et ne sont pas étrangères au statut social et politique de la langue. Les études occitanes souffrent de plus d'une segmentation des pratiques et des réflexions. Spécifiquement, la circulation n'est pas facile entre la description des variétés dialectales de l'occitan (même pénétrée de la dignité de la langue) et la codification (même bien informée de la réalité dialectale). Notre propos n'est pas ici d'analyser les raisons de ces dysfonctionnements, mais de montrer par l'exemple que la codification graphique de l'occitan a une utilité scientifique et descriptive, en plus de sa pertinence pratique première (écrire, utiliser et enseigner la langue)[3]. Pour ce faire, nous présenterons ici la logique de la codification graphique dite « classique » de l'occitan avant de montrer comment une lemmatisation fondée sur cette graphie est un précieux instrument d'organisation interne de la masse de données lexicales et morphologiques occitanes que constitue le Thesaurus Occitan (THESOC, cf. Dalbera 1998 et pour la partie en ligne : http://thesaurus.unice.fr/).

Abstract : Too few scholars are dedicated to Occitan studies. There are a lot of reasons for such a situation, among which the social and political status of the language is not the least. Occitan studies also are affected by a split in the research attitudes and conceptions.

1. Université de Toulouse II – Le Mirail & Laboratoire mixte CNRS / UTM, UMR 5263.

2. Laboratoire mixte CNRS / UNSA, UMR 7320.

[3] La pertinence de la codification pour la recherche linguistique sur l'occitan est développée dans Sauzet 2002.

Specifically, there is no regular continuity between works describing dialectal varieties and works contributing to corpus planning, even when the former ones don't undervalue the language status and the latter ones don't ignore dialect complexity. We shall not try here to analyse the reasons for this difficult cooperation. We only want make it clear, on the basis of a few examples, that the orthographic codification of Occitan not only has a practical relevance (allowing to write, use and teach the language), but also is a valuable scientific and descriptive tool[4]. In order to do so, we shall first explain the principles of the so-called "classical" Occitan orthography and we shall then indicate how lemmatizing on the basis of this notation strongly helps organize, from inside the language, the huge wealth of lexical and morphological data included in Thesaurus Occitan (THESOC, cf. Dalbera 1998 and for an online sample : http://thesaurus.unice.fr/).

Introduction

On ne peut en faisant de la linguistique ou de la dialectologie occitane échapper à la situation sociolinguistique de cette langue. Le rappel de cette évidence ne manquera pas de recueillir l'assentiment teinté d'ironie. Il recèle pourtant une ambiguïté qu'il est utile d'expliciter.

On peut l'entendre comme un constat. La façon dont on fait de la linguistique ou de la dialectologie est nécessairement marquée par la situation sociolinguistique de la langue étudiée. Le linguiste est, qu'il le veuille ou non, un acteur du jeu de dominance et de concurrence où sont impliquées les langues en domaine occitan. On peut l'entendre aussi comme une injonction. On ne saurait se pencher sur l'occitan sans tenir compte activement de la situation de cette langue, sans reconnaître l'occitan comme langue dominée.

À chaque interprétation s'attache potentiellement un type de démarche biaisée. L'effet non corrigé de la situation sociolinguistique sur l'observateur conduit à pratiquer sans distance une linguistique diglossique. Le linguiste inscrit son discours dans la langue dominante, le français, qui est

[4] Cf. Sauzet 2002 for discussion.

notamment sa langue de travail, depuis laquelle il observe l'occitan, qu'il appellera éventuellement 'patois'. Produit dans ces conditions le discours scientifique sur l'occitan légitime la relégation sociale et culturelle de la langue, quels que puissent être par ailleurs les protestations, démonstrations ou sentiments réels d'attachement pour la langue étudiée. Quelles que soient aussi les qualités du travail descriptif et analytique produit. Une linguistique qui utilise par réalisme le terme de 'patois', comme les locuteurs eux-mêmes, n'adopte pas une attitude neutre et ne se situe pas dans quelque degré zéro de la glossonymie. En la reprenant à son compte, il valide la désignation courante de son autorité de spécialiste du langage. De plus, en reprenant comme telle la désignation ordinaire des témoins, le linguiste leur attribue exclusivement cette langue dont ils garantissent la nomination[5] : cette langue est la leur, pas la sienne, même s'il la parle aussi. Nommée 'patois', la langue décrite ne saurait être celle du linguiste.

Dans la logique de la situation diglossique, le linguiste est universitaire et francophone[6]. Il peut aussi être étranger, mais il a alors acquis le français comme langue d'enquête. Le linguiste peut bien sûr être en outre locuteur d'occitan. Éventuellement il peut biographiquement avoir été d'abord occitanophone. Il reste que la métalangue de son activité n'est pas l'occitan mais en général le français. La *Revue des langues romanes* en ses débuts accueillait l'écriture littéraire en langue d'oc à côté des travaux de romanistique, mais ces derniers étaient exclusivement en français. Même Mistral en rédigeant le *Tresor dóu Felibrige* (d'une grande modernité lexicographique, qui, combinée avec

[5] Un pas qu'il faut se garder de franchir consiste à considérer que 'patois' étant un terme couramment utilisé par les locuteurs serait un 'autoglossonyme'. Pour être assumé, 'patois' n'en reste pas moins historiquement induit sinon imposé. Cette historicité est soulignée par la perception différente du terme hors des frontières politiques françaises, comme par sa non-implantation en pays niçard sur laquelle insiste un de nos relecteurs.

[6] Il s'agit de la situation, très largement majoritaire, des parlers occitans en territoire français. La situation sociolinguistique est notablement différente (et pas seulement parce que la langue haute est autre que le français) dans les Vallées occitane italiennes ou le Val d'Aran espagnol.

les carences de l'université occitane, le rend pertinent et indispensable encore aujourd'hui) utilise le français comme métalangue. La *Gramatica occitana* d'Alibert de 1935 (Alibert 1976) est bien en occitan, mais son dictionnaire (Alibert 1966) est un dictionnaire occitan-français. On peut opposer cette pratique occitane à la pratique catalane dont la renaissance linguistique a comporté d'emblée un usage métalinguistique du catalan (cf. Schlieben-Lange 1971, 41-42).

Ne rien faire, laisser agir les forces en présence, c'est conforter la domination. Le constat explicité du poids de la diglossie sur le linguiste conduit à la deuxième lecture évoquée plus haut, la lecture injonctive. Puisqu'on ne saurait être neutre, choisissons notre camp ! Choisissons éventuellement de pratiquer une linguistique de contestation de la diglossie. Cela pourra se marquer d'abord par la revendication d'un champ sociolinguistique spécifique, par une manipulation volontariste des dénominations linguistiques ('patois' étant repéré comme désignation diglossique de l'occitan), éventuellement aussi par un recours d'ampleur variable à l'occitan comme métalangue. Si la première démarche peut être entachée de complicité avec la diglossie, la dernière démarche peut mener à un aveuglement volontariste qui, soit n'aborde l'occitan que du point de vue sociolinguistique, comme théâtre de ce qu'on décrira volontiers comme un conflit linguistique (au détriment de la linguistique interne), soit s'épuise dans des constructions normatives, dont la normalité conquise se relie difficilement aux faits obstinément diglossiques.

Ce qui vient d'être décrit, ce sont les idéalisations de tendances auxquelles les œuvres des linguistes réels cèdent plus ou moins volontiers, ou résistent avec plus ou moins de succès. Elles aideront ici à comprendre comment se pose la question de la lemmatisation dans le domaine occitan. La lemmatisation est en effet une pratique métalinguistique interne : son absence prolonge l'absence d'usage métalinguistique de l'occitan, et sa mise en place conduit à s'interroger en général sur les formes de l'usage métalinguistique en domaine occitan, donc sur des formes linguistiques qui relèvent de la norme, du standard, de la codification.

Typiquement, l'*Atlas Linguistique de la France* (ALF, Gilliéron & Edmont 1902-10) dans ses cartes comme dans son index, ne connaît à peu près (nous reviendrons plus loin sur l'approximation) que les termes français dans lesquels sont posées les questions et les formes phonétiques des réponses. À l'opposé, l'*Atlas linguistique de Catalogne* donne dans les intitulés des cartes la ou les formes orthographiques catalanes (en plus des équivalents castillans, italiens et français). Les atlas régionaux, réalisés dans le cadre du *Nouvel Atlas Linguistique de la France* (NALF), présentent, malgré quelques hésitations, quelques pas dans le sens d'un usage d'une lemmatisation interne de l'occitan. Elle peut prendre la forme de la notation phonétique d'une forme moyenne. Ainsi, dans les carnets d'enquête de l'*Atlas Linguistique du Languedoc Occidental* (ALLOc, Ravier 1978-93), il arrive que la forme typique attendue soit ainsi écrite, en notation phonétique dite 'Rousselot' : ex. [kalim'as], qui glose « chaleur étouffante » et indique le type de réponse attendu. Les lemmes sont indispensables, et des formes graphiques bien commodes pour noter les données négatives qu'introduit l'*Atlas Linguistique de la Gascogne* (ALG, Séguy 1954-73) : il est difficile de citer autrement que dans une notation conventionnelle une forme absente. Or l'orthographe est précisément une convention reçue. L'ALG utilise quelquefois des formes orthographiques dans cet usage (quelquefois aussi la base latine de la forme rejetée), l'ALLOc et l'ALLOr (*Atlas Linguistique du Languedoc Oriental*, Boisgontier 1981-86) le font assez systématiquement.

La diglossie, spécifiquement le bilinguisme diglossique, est au cœur des atlas linguistiques, de l'ALF comme des atlas régionaux, dans la mesure où le français est la langue d'enquête. En domaine d'oc, la traduction est parfois revendiquée non comme une contrainte mais comme un avantage méthodologique, puisqu'elle est censée éviter la contamination des formes locales par un modèle que l'enquêteur ne manquerait pas de suggérer. L'inconvénient inverse, du poids du modèle français sur les formes locales, outre qu'il est réputé plus facilement repérable, a été corrigé dans les nouveaux atlas qui d'une part, dans l'enquête, tendent à pratiquer plus la définition

que la nomination directe, et d'autre part ont renoncé à la doctrine de la première réponse de règle dans l'ALF. Les enquêteurs de la seconde vague d'atlas vont jusqu'à pratiquer, en désespoir de spontanéité et en le mentionnant, la suggestion de formes occitanes.

Il y a donc toujours une forme française à laquelle rattacher une carte d'atlas linguistique. En face, une multiplicité de formes occitanes. Dans l'ALF, c'est toujours (dans l'atlas lui-même, pas dans les travaux qui en sont issus) une multiplicité brute, autant de formes que de points d'enquête[7]. Certains atlas régionaux font le choix de présentations par aires dégagées qui construisent quelque fois une quasi lemmatisation (ALG, ALLOc, ALLOr, etc.).

Ce n'est pas ici le lieu de discuter si une autre méthodologie, qui installe l'occitan comme métalangue au départ d'une entreprise d'atlas linguistique, aurait été possible. En revanche, la prise en compte réaliste de la façon dont les atlas ont été construits et du poids que la situation sociolinguistique a pesé sur leur construction, n'interdit pas de recourir à des outils dont la logique est celle de la contestation de la diglossie pour mener à bien leur exploitation. Toute orthographe, toute graphie conçue pour un usage social est, à quelque degré, inscrit dans une telle contestation de la diglossie[8].

Il importe toutefois de relever que, de même que des travaux inscrits dans le consentement à la diglossie (voire à la disparition de la langue étudiée) peuvent néanmoins constituer un apport scientifique irrécusable (et sur certains point tirer des bénéfices de la forme de démarche induite par l'acceptation de la situation sociolinguistique, comme le recours à la traduction), de même le recours à des formes orthographiques est aussi un outil scientifique utile, indépendamment de son usage social possible.

[7] Dans l'index de l'ALF on trouve toutefois une forme de lemmatisation par allègement des diacritiques sur les formes phonétiques et par l'enregistrement de type français régionaux qui sont des calques des formes locales éventuellement occitanes.

[8] Ce qui ne signifie pas que tout usager d'une notation orthographique doive afficher un militantisme linguistique radical.

Considérons l'exemple classique des formes du nom du « chien » en domaine occitan (cf. Dauzat 1929). En recourant à des formes orthographiques occitanes, on énonce d'emblée qu'il y a quatre aires principales, celles de *can*, de *chin*, de *chen*, de *gos*. Ces désignations sont disponibles de manière à peu près évidente et consensuelle pour toute personne qui sait manier la graphie classique de l'occitan. Un résultat approchant serait obtenu avec la graphie mistralienne (on noterait seulement *gous* la dernière forme et il faudrait substituer plusieurs notations à *can*). Sans recours aux formes orthographiques, une solution classique consiste à remonter aux types latins. Dans ce cas spécifique (en s'en tenant aux hypothèses couramment admises) il faudrait poser trois fois CĂNEM à la source des formes concernées et une forme $^{?}$°(SE)GUSIU(M) ou autre pour *gos*. Il faudrait en fait trancher et poser que, si *can* remonte bien à CANEM, *chin* et *chen* remontent à des prototypes français et francoprovençaux dont on voit mal comment les noter à leur tour autrement qu'orthographiquement. Sans compter qu'il est gênant de trancher (même en faisant la part de la dimension méthodologique de la décision) sur les formes sources de *chin* et de *chen* pour construire la discussion de cette forme source. Sans compter non plus que ces formes sources devront aussi être notées orthographiquement.

Les lemmes sont utiles pour manipuler la langue quand il s'agit, comme on vient de le voir, de discuter de la variation et de l'histoire des mots. Dans l'article évoqué comme ailleurs dans son œuvre, Dauzat, par ailleurs archétype du linguiste usant du terme « patois » et grand pourfendeur de l'enseignement de l'occitan, utilise volontiers les formes médiévales en guise de lemme : l'article cité s'ouvre par la phrase « Voilà longtemps que l'équation AQUA > prov. *aiga* a donné de la tablature aux romanistes. » L'abréviation *prov*. « provençal » renvoie à l'occitan et spécifiquement à l'occitan médiéval. La forme de l'ancien occitan sert à discuter du destin de ce mot à travers l'histoire et l'espace de la langue.

Les lemmes sont aussi et d'abord indispensables à toute entreprise lexicographique : tout dictionnaire occitan dont l'ambition dépasse une microrégion se pose la question de sa

lemmatisation. C'est vrai du *Tresor dóu Felibrige*, du dictionnaire d'Alibert, de celui de Palay (Palay 1932-33) même si les réponses divergent. C'est vrai aussi de l'entreprise du THESOC qui a intégré dès le début la lemmatisation dans son système de saisie. Les lemmes et, de manière à la fois plus générale (parce qu'il ne s'agit pas de formes isolées) et plus spécifique (parce qu'il s'agit d'un type de lemme particulier), la notation orthographique sont des outils indispensables pour transcrire de manière large des textes oraux et entreprendre des travaux de syntaxe.

1.Propriétés des graphies classique et mistralienne

Étant admis le besoin de lemmes et de lemmes orthographiques, quel choix faire parmi les normes graphiques possibles de l'occitan (ou de la langue d'oc).

Il faut tout d'abord fixer les désignations. D'un côté, on parle ou l'on entend parler de graphie mistralienne, roumanillienne, félibréenne, provençale… De l'autre, de graphie classique, occitane, normalisée, alibertine… Nous retenons « mistralienne », malgré les hésitations initiales de Mistral en fait de graphie et l'influence de Roumanille sur lui sur ce point, parce que l'adoption de cette notation par Mistral dans son œuvre[9] et dans son dictionnaire en sont la référence et l'appui majeurs. « Félibréenne » ne convient pas parce que le Félibrige utilise plusieurs graphies, « provençale » encore moins puisqu'on peut écrire, qu'on a écrit et qu'on écrit le dialecte provençal avec d'autres graphies et que d'autres dialectes peuvent s'écrire en graphie mistralienne. Nous retenons « classique » pour l'autre graphie envisagée parce que sa référence centrale est la réactivation des notations médiévales, régularisées et complétées de diacritiques. « Alibertine » n'est pas faux, mais isole le travail décisif du linguiste éponyme de la continuité où il s'inscrit. « Occitane » suggère qu'il n'y a qu'une graphie possible pour

9 Ce qui n'implique pas que l'œuvre de Mistral ne soit pas graphiquement transposable bien sûr.

l'occitan ou langue d'oc (alors qu'elle peut comme toute langue s'écrire de mille manières) ou que n'est occitan que ce qui s'écrit avec cette graphie (alors que ce n'est pas le vêtement graphique qui fait la langue d'oc ou occitan, et que ce choix peut tout au plus contribuer à rendre son unité plus visible).

Les graphies 'mistralienne' et 'classique' ont en commun d'être des orthographes autonomes, liées à une entreprise de restauration ou de renaissance littéraire et linguistique[10]. L'une comme l'autre sont nées dans le Félibrige : la mistralienne en est la « loi » originelle, la classique un développement critique. Elles sont de ce fait liées l'une à l'autre à plus d'un titre : la graphie classique se construit à partir de la graphie mistralienne qu'elle entend améliorer ou accomplir, la graphie classique retient aussi des solutions qui avaient été envisagée dans la gestation de la graphie mistralienne (*-a* final, notation de marques de pluriel localement muettes)[11].

1.1 Élément communs

On peut référer les éléments communs à un même principe de **sobriété** qui règne dans des orthographes ayant en vue l'usage régulier (et non l'utilisation occasionnelle)[12]. Elle s'oppose en cela à ce qu'on peut appeler « graphies oralisantes », c'est à dire des graphies orientées vers la restitution de l'effet sonore par le transfert des conventions graphique du français. Il est clair que la notation des manuscrits de l'Abbé Fabre par exemple vise avant tout à faire entendre la langue qu'elle note.

[10] J. Taupiac insiste justement sur les rationalisations qu'apporte Mistral à Honnorat et que conservera Alibert, même si certains choix d'Honnorat préfiguraient ceux de la graphie classique (Taupiac 1980).

[11] Pour une présentation de la graphie classique voir en particulier Lafont 1971 et 1972. Pour l'histoire de la codification de l'occitan au début du XXe siècle, voir Kremnitz 1974.

[12] La formule suivante que Ronjat applique à "l'ortografe félibréenne" vaut tout aussi bien pour la graphie classique: "(cette orthographe) rejette en principe tout **signe** fisiologiquement ou psicologiquement **superflu**." (graphie du français et mise en relief de l'auteur) (Ronjat 1930-37 1 §45, 81).

Cette sobriété ou cette économie s'exprime dans les choix suivants :

- notation légère et unifiée des diphtongues avec *-u* ou *-i* comme second élément : *au, ai*… *paure, paire* dans les deux graphies contre des notations 'aou, âou, aù, âi, aï, ay…' dans les usages oralisants,

- notation légère et unifiée des affriquées : *chato* en graphie mistralienne (désormais « g.m. »), *chata* en classique (désormais « g.c. ») (et non 'tchato, tchiato, tsato…'), *jamai* ('djamâï, djiamaï, djhamaï, dzamai…'),

- non-notation de l'accent tonique non marqué : *cabro* (g.m.), *cabra* (g.c.) ('câbro, câbra…'), *dire* ('dîré, dirë….'),

- rejet des notations ornementales : *filousoufio* (g.m.), *filosofia* (g.c.),

- rejet des diacritiques redondants (voir ci-dessus diphtongues et notation de la place de l'accent) : *e* (sans accent) note [e] *que se netege e se seque*, dans les deux graphies.

Outre leur sobriété, les deux orthographes de l'occitan se rejoignent dans la notation de certains faits de variation dialectale. Les résultats d'évolutions lourdes et anciennes sont notés. C'est le cas par exemple et typiquement :

- de la palatalisation nord occitane : *chabro, cabro* en g.m., *chabra, cabra* en g.c.,

- du traitement gascon du *-f-* latin : *hilho, filho~fiho* en g.m., *hilha, filha* en g.c.,

- de l'article pluriel caractéristique du provençal moderne : *li biòu, lei biòu* en face de *lous biòus* en g.m., *lei* (= [li] ou [lej]) *buòus* en face de *los buòus* en g.c.

1.2 Différenciation arbitraire

En revanche, les deux graphies se différencient par quelques choix arbitraires. Dans ces choix, la graphie mistralienne privilégie en général la communauté avec le français alors que la graphie classique privilégie la continuité avec les usages médiévaux spécifiques à l'occitan. On oppose ainsi :

- la notation de la nasale palatale qui est *gn* en g.m. et *nh* en g.c. : *vigno* vs *vinha*,

- la notation de la latérale palatale historique qui est *-i-* ou *-h-* en g.m. pour les parlers qui remplacent cette palatale par [j], et *-lh-* pour les autres, alors qu'elle est systématiquement *-lh-* en g.c. : *paio, fiho* ou *palho, filho* selon les parlers en g.m., *palha, filha* partout en g.c.,

- la notation des sons [ɔ] et [u] est *o, ou* respectivement en g.m. mais *ò, o* en g.c. : *lou drolle* en g.m. pour *lo dròlle* en g.c..

Dans le premier cas, la g.m. adopte la solution franco-italienne, tandis que la g.c. recourt à une notation fréquente au moyen âge en domaine d'oc (d'où elle a passé au portugais). *nh, lh* font système pour la notation des palatales avec *ch*, et *sh, th* utilisés en gascon.

Dans le cas de *o, ou* ou *ò, o*, la graphie médiévale était insuffisante, donnant à « o » deux valeurs [ɔ] et [ɔ] ou [ɔ] et [u] selon les époques. Mistral a recours à la notation française de [u], « ou » et attribue ainsi à « o » la seule valeur de [ɔ]. Alibert aligne le système de *ò, o* sur celui de *è, e* qui notent [ɛ] et [e] respectivement dans les deux graphies. Les alternances en morphologie de *ò* avec *o* et de *è* avec *e* sont de fait strictement parallèles : *jòga, jogar* comme *lèva, levar*… (en g.m. : *jogo, jouga, lèvo, leva*).

1.3 Différenciation par le caractère englobant

Les deux graphies ou les deux orthographes (si l'on entend par orthographe un système graphique fixé à vocation d'usage social, une graphie pouvant n'être que technique ou didactique) divergent surtout par leur degré d'intégration graphique ou d'univocité. La graphie mistralienne note généralement plus près de la variation dans un certain nombre de cas, alors que la graphie classique est plus englobante[13].

[13] Cette différence entre les deux graphies est soulignée par exemple dans Teulat 1980.

Cette différence est manifeste autour des faits suivants :

- notation des produits de *-a* post-tonique protoroman : la g.c. note exclusivement *-a* qui recouvre des réalisations [o], [ɔ], [a], [ə], [u] tandis que la g.m. note *-o* pour [o], [ɔ], *-a*[14], *-e*, *-ou*,

- notation d'évolutions vocaliques diverses 'récentes' (post médiévales) par la g.m. alors que la g.c. les laisse de côté : diphtongaison de /ɔ/ *pòrc* en g.c. pour *porc*, *pouorc*, *pouerc*, *pouarc*, *puerc*… en g.m., labialisation de /a/ en diverses positions : notation unique *campana* en g.c. pour *campano*, *campono*, *compano*, *compono* en g.m. ; la g.m. note l'évolution éventuelle des diphtongues atones *eigueto* à côté d'*aigo* dans les parlers qui connaissent ce phénomène alors que la g.c. note comme ailleurs *aiga* et *aigueta*… ; la g.c. note la diphtongaison conditionnée que les textes médiévaux montrent en train de se développer, mais elle réduit le nombre des types notés : *-uò-*, *-uè-*, *-ue-* pour le produit de *-ò-* en contexte palatal (*nuèch* ~ *nuèit*, *nuech*, *nuòch*), *-iè-*, *-ie-* pour le produit de *-è-* dans le même contexte (*vièlha*, *vielha*) les réductions ou différenciation ultérieures de la diphtongue pouvant donner par exemple des réalisations [ɲˈø] ou [nˈet] ne sont pas notées spécifiquement[15],

[14] Le timbre [a] de la finale, tel qu'on le trouve à Montpellier ou à Nice, impliquerait dans la logique graphique mistralienne la notation systématique de l'accent tonique pénultième puisque *-a* final est censé être accentué : *canta* [kantˈa] implique de noter *cànta* pour [kˈanta]. En fait dans les parlers qui réalisent [a] atone final, on note en g.m. l'accent sur [a] tonique final: *canta*, *cantà* (Nice ou Montpellier) en face de *canto*, *canta* (Maillane ou Toulouse) (Ronjat §43, 80). En gra. clas., on a dans ce cas uniformément *canta* et *cantar*.

[15] Du moins ne le sont pas en principe, car on trouve des usages de la graphie classique qui notent par exemple des formes comme *nèit* ou *neit*... Remarquons en passant que la différenciation de *-uè-* et de *-ue-*, de *-iè-* et de *-ie-* que retient la gra. clas. pourrait avantageusement être abandonnée dans une entreprise lexicographique sinon dans l'usage. Cette notation qui ne tend qu'à rendre des différences dialectales et non des oppositions internes à un parler est contraire à l'esprit général de la gra. clas. Un point d'application particulier est la graphie du suffixe *-ier* < ARIU où *-ièr* languedocien s'oppose à *-ier* provençal mais sans pertinence phonologique dans aucun des dialectes. On écrira avantageusemenet *-ier* partout (mais *-èr*, *-èir* en gascon), de même que *nuech* ~ *nueit* et *vielh*.

- notation des consonnes latentes : elle est systématique dans la gra. clas. alors qu'elle est exceptionnelle dans la g.m. :

- la graphie classique écrit partout *parlat, prat, sec, dich ~ dit* que les occlusives finales soient réalisées (comme en gascon ou en languedocien) ou non (comme en provençal ou nord-occitan), la g.m. note selon la réalisation locale *a canta* ou *a cantat, se,* ou *sec, di* ou *dich* ou *dit ;* elle note néanmoins toujours *prat* réalisé [prˈa] ~ [prˈat],

- la graphie classique écrit *canton, segur* que l'*-n* ou l'*-r* finale se prononce comme en Provence ou qu'elle soit muette comme c'est le cas plus à l'ouest ; la g.m. note *cantoun* ou *cantou, segur* ou *segu.*

- notation des neutralisations consonantiques finales : la graphie classique ne les note pas *cantam, codonh* qu'*-m* et *-nh* gardent leur valeur respective de [m], [ɲ] ou se confondent en [-n] , *trabalh,* que *-lh* soit [ʎ] ou [j] ou [l] alors que la g.m. note *cantan* ou *cantam, coudounh*[16] ou *coudoun, trabai (travai), trabalh* ou *trabalh,*

- notation du bétacisme : un grand tiers sud-ouest du domaine occitan ignore totalement la labiodentales sonore, la g.c. note pourtant systématiquement « v » dans tous les dialectes et écrit toujours *vinha, parlava* tandis que la g.m. note *vigno* et *bigno, parlavo* et *parlabo* ; *parlava* ou *lavar* supportent aussi les réalisations gasconnes où « v » note [w] alors que ces formes en g.m. demandent une notation spécifique : *parlauo, laua.*

On peut synthétiser la différence essentielle entre les deux graphies en disant que la graphie mistralienne est plus phonétique et la graphie classique plus phonologique. Il en résulte que les deux orthographes sont plus voisines pour un parler où les processus phonologiques sont moins complexes et plus différenciées pour un parler à la phonologie plus riche. Très caractéristiquement, c'est pour un parler comme le provençal rhodanien que les notations des deux graphies sont maximalement distantes. C'est sans doute une des racines de la vivacité polémique que revêt quelque fois le débat graphique

[16] Cf. TdF s.v. *coudoun.* D'ailleurs Mistral, pour une raison qui nous échappe, note *bagn* et *banh,* à côté de *ban* (pour gra. clas. *banh*). Ronjat lui ne note qu'en *-gn* : *bagn* les réalisations palatales (Ronjat §52, 95).

occitan, encore qu'on puisse aussi penser que cette vivacité est un trait commun à tous les débats de ce type. Que l'on songe aux débats qui secouent l'Allemagne au moment où nous rédigeons cet article pour l'usage du ß ou à ceux que soulève la moindre discussion d'un circonflexe français !

Le provençal rhodanien amuit les obstruantes finales. La g.m. ne note qu'une partie de ces consonnes devenues latentes (parce qu'on les retrouve dans la dérivation). Elle note ainsi pour le provençal rhodanien *cat*, *loup* (comme *cat*, *lop* en g.c.) mais *ro* ou *fiò*… (*òc*, *fuòc* en g.c. pour tous les parlers). La g.m. note, dans les parlers qui amuïssent les occlusives finales, *lou roustit* parce qu'il s'agit d'un nom, mais *es rousti* parce que c'est un participe passé (en g.c. on a uniformément, dans les parlers amuissants et dans les autres : *lo rostit*, *es rostit*).

On voit donc que dans de nombreux cas où la graphie mistralienne doit épouser la variation, la graphie classique propose d'emblée une forme unique. Il reste que la graphie classique enregistre certains processus évolutifs divergents. Ce sont les processus caractéristiques des grands ensembles dialectaux soit :

- la palatalisation qui oppose le nord occitan au sud depuis l'origine de la langue : *chabra*, *jau* au nord pour *cabra*, *gal* au sud,

- la débuccalisation de l'*f* en gascon, *hilh*, *hlor* vs *filh*, *flor* ainsi que d'autres processus typiques de l'aire gasconne : chute de l'*n* intervocalique : *plea* pour *plena*, traitement spécifique de -LL- géminé latin : *aquera bèra vedèra*, *aqueth bèth vedèth* pour *aquel(a) bèl(a) vedèl(a)*.

- les deux traitements du groupe -CT- latin, le sud-ouest conservant [-jt-] *-it-* que l'est et le nord (sauf une zone d'Auvergne et du Croissant) remplacent par une palatale *-ch-* : *faita* vs *facha* de FACTA etc.

- la vocalisation de l'*-l* final qui oppose le centre du domaine à toute la périphérie occitane : *ostal* vs *ostau*.

La gra. clas. note aussi les processus qui affectent la composition segmentale des mots : métathèses *crompar* en face de *comprar* ; *craba* en face de *cabra*, *cramba* en face de *cambra*,

merulhier pour *melhurier*… les aphérèses : ‘*nar* pour *anar*, les prothèses *arriu* pour *riu*…

La graphie classique réunit donc assez largement les formes occitanes issues d’un même étymon latin, mais pas totalement. Tant pour l’usage didactique ou pratique que pour l’usage scientifique de la lemmatisation, il peut être utile de disposer d’une forme de référence au-delà de ce que la graphie établit d’elle-même. Il s’agit là d’une lemmatisation lexicale par opposition à la lemmatisation graphique. Cette opération demande des choix et des principes sur lesquels ces choix reposent.

Aucune des deux graphies présentées n’implique par elle-même de choix particulier concernant cette lemmatisation lexicale. Chacune d’elle est néanmoins associée par l’histoire de son développement à une logique propre en ce domaine. La g.m. s’accompagne naturellement d’une promotion systématique de la forme d’occitan utilisée par Mistral dans ses œuvres, le provençal rhodanien, nommé dans sa fonction de langue référence ‘provençal littéraire’. C’est ce que pratique Mistral lui-même dans son dictionnaire : les entrées s’ouvrent systématiquement sur la forme rhodanienne que suivent quand elles en diffèrent les variantes attestées des autres dialectes. J. Ronjat rejoint cette pratique : la *Grammaire istorique*, pour chaque phénomène étudié et en particulier pour analyser la morphologie, présente d’abord le provençal littéraire, note ensuite les traits du rhodanien (dit « rhodanien populaire ») que ne retient pas la langue littéraire, puis envisage les autres dialectes en commençant par ceux qui sont « voisins du provençal littéraire » et en s’en éloignant progressivement.

Les utilisateurs de la g.c. (mais ceux de la g.m. aussi) ont des positions très contrastées sur la norme linguistique qui peut, doit ou ne doit pas prolonger la fixation graphique. Il reste que l’entreprise de construction de la graphie dite classique est depuis le début fortement articulée sur la langue médiévale dont elle reprend fondamentalement les notations. Dans cette logique, la lemmatisation se fait en direction des formes les plus proches de la langue médiévale. La graphie classique est aussi née d’une volonté de prise en compte globale de l’espace d’oc. Il y a

plusieurs façons de promouvoir parmi des parlers ou des usages une forme à vocation référentielle (que ce soit à des fins sociales ou techniques). On peut assumer un choix linguistiquement arbitraire et s'en remettre à une légitimation externe au système. C'est la logique de la lemmatisation sur le provençal littéraire. Ce qui valide le caractère référentiel du provençal rhodanien, c'est l'œuvre de Mistral et des félibres qui partagent son usage linguistique. Ce ne sont en aucun cas les propriétés intrinsèques de ce dialecte. Dans la logique de la g.c., un parler peut être considéré comme plus apte à jouer le rôle de référence ou de standard (sans encore une fois préjuger du rôle qu'on assigne à cette forme linguistique) à cause de ses propriétés. Aura typiquement vocation référentielle, un parler qui présente moins d'innovations et surtout moins d'innovations spécifiques. Les parlers languedociens sont centraux dans l'ensemble occitan au sens qu'ils ont un confront avec à peu près[17] chacun des autres dialectes. Pour la plupart des faits évolutifs que note la graphie classique, le languedocien est du côté de la conservation :

- pas de palatalisation nord occitane des vélaires latines[18],
- pas de débuccalisation gasconne du « F » latin
- pas de vocalisation nord-occitane, gasconne ou provençale d'*l* final,
- conservation intégrale (au niveau morphémique et donc écrit) du pluriel sigmatique.

Cela ne signifie pas que le languedocien soit systématiquement « conservateur ». Les parlers languedociens sont sûrement souvent innovateurs en matière morphologique ou lexicale. Dans le domaine phonologique : la chute d'[-n] ou [-r] finaux sont des innovations par rapport à l'état roman commun ou médiéval, le bétacisme, originel en gascon, est relativement récent en languedocien, post médiéval en tous cas. Comme ces

[17] Le confront languedocien-vivaro-alpin est réduit sinon théorique.

[18] La palatalisation n'est une "innovation" que relativement à l'état latin ou protoroman. Dans l'histoire de la langue elle est attestée dès les premiers textes, comme bien des traits gascons et à la différence des principaux traits qui caractérisent le provençal.

faits ne sont pas notés en g.c. ils n'affectent pas l'aptitude du languedocien à fournir des lemmes panoccitans.

Louis Alibert n'avait pas particulièrement en vue un rôle référentiel panoccitan du languedocien en rédigeant sa *Gramatica* puis son dictionnaire. Il se trouve que son œuvre est compatible avec une architecture normative à deux niveaux[19]. À un premier niveau, les grands ensembles dialectaux développent des formes de standards si l'on se place du point de vue social ou des systèmes de lemmatisation essentiellement graphique si l'on prend les choses techniquement. À un second niveau, une forme de languedocien joue le rôle de standard commun pan occitan pour qui juge pertinent l'usage d'une telle forme et peut jouer rôle de lemme dans une entreprise lexicographique ou lexicologique panoccitane comme le THESOC.

2. La lemmatisation au sein du Thesoc

2.1 Lemmatisation et indexation :

Grâce à la participation de Guylaine Brun-Trigaud au projet d'indexation des atlas linguistiques régionaux[20], les tâches de lemmatisation et de recherches étymologiques qui lui ont été confiées au sein de l'équipe du Thesoc ont bénéficié de l'expérience ainsi acquise. Cependant, le travail de lemmatisation est un peu différent du travail d'indexation, puisque dans le cas des atlas, il fallait rendre compte entièrement du contenu des cartes, en effaçant toute la variation phonétique au profit d'un lemme, quitte parfois à laisser des formes en suspens, avec une simple translittération.

L'idée des index d'Atlas est assez ancienne, puisque J. Gilliéron et E. Edmont ont fait paraître un pour l'ALF dès 1912. Elle fut reprise par P. Gardette dans *Commentaires et Index* (1976) pour l'*Atlas Linguistique du Lyonnais* qui avoue toute la difficulté de l'entreprise dans un domaine où il n'y a pas de

[19] Sur l'architecture générale de la norme classique occitane on peut se référer à la synthèse de Dominique Sumien (Sumien 2006).

[20] Cf. Brun-Trigaud, 2003a et 2003b.

tradition orthographique officielle (p. VII-VIII) : la majeure partie de l'ouvrage est consacré aux commentaires des cartes avec renvois au *Französische Etymologische Worterbuch* (FEW) et l'index contient massivement des formes en graphie phonétique.

Plus tard, l'informatique ouvrit de nouvelles perspectives et G. Taverdet compila et publia les *Index* de trois atlas (*Atlas Linguistique de la Bourgogne* en 1988, *Atlas Linguistique de la Champagne et de la Brie* en 1989 et *Atlas Linguistique du Centre* en 1993) en utilisant le système graphique du français, à quelques rares exceptions pour les formes faisant difficulté laissées en graphie phonétique.

Entre temps, en 1991, un projet d'indexation portant sur l'ensemble des atlas régionaux avait été lancé par le GDR des Atlas : l'idée était de constituer un index pour chaque atlas, avec possibilité d'introduire des bases étymologiques, le tout devant être regroupé dans une base nationale. Deux logiciels plus tard et après la dissolution du GDR des Atlas, l'indexation a poursuivi son chemin par des procédés différents. C. Dondaine a fait paraître en 2002 un *Trésor étymologique des mots de la Franche-Comté, d'après l'ALFC*. L'index alphabétique des formes regroupe la totalité des formes dialectales, typisées avec une graphie spéciale, suivies d'un renvoi au FEW et au *Glossaire des Patois Suisses Romans* ... En 2010, F. Carton et Alain Dawson ont publié l'index de l'*Atlas Linguistique Picard*, en suivant les principes adoptés pour le projet initial du GDR des Atlas.

2.2 Le rôle de la lemmatisation dans le Thesoc :

La lemmatisation joue un rôle important dans le Thesaurus Occitan et elle intervient à plusieurs niveaux.

L'architecture de la base de données en ce qui concerne l'implémentation des données fonctionne avec quatre niveaux :

1) la forme phonétique est saisie en transcription API, induisant une normalisation par rapport à la notation phonétique des atlas (= champ phonique).

2) la graphie phonologisante permet de restituer en graphie de type mistralien la forme phonétique (elle « traduit » la phonétique pour un public souvent peu familier des alphabets phonétiques : par ex. [ysel'ũ] « moineau » (La Javie, ALP pt 69) sera transcrit *usseloun* (= champ graphie phonologisante).

Ph. Dalbera et D. Strazzabosco ont d'ailleurs mis au point un premier algorithme permettant de réaliser automatiquement cette transcription qui fonctionne parfaitement bien pour les aires provençales et languedociennes, mais qui éprouve ses limites dans les autres aires, d'où des aménagements à venir sur cette fonction bien utile.

3) le lemme genre de « chapeau » à « plusieurs casquettes » : sa fonction principale de « chapeau » est de permettre de rassembler sous une même forme graphique de type alibertin (puisée dans le *Dictionnaire Occitan Français selon les parlers languedociens* (DOF)) les différentes variantes phonétiques d'un même terme : ex. pour le terme « chèvre », parmi les 605 réponses recueillies actuellement dans les atlas, on relève plus de 40 variantes phonétiques différentes [k'abrɔ, ʃˈæbr, ʃjˈɛb, ʦˈabrœ, ʧˈurɔ], etc. que l'on peut regrouper sous *cabra* (= champ lemme).

Ses différentes « casquettes » sont les utilisations que l'on peut en tirer : d'une part, ce regroupement permet de dessiner des cartes à aires lexicales où les lemmes sont représentés par des cercles colorés donnant la répartition géographique des termes recueillis pour une notion donnée, d'autre part, cela autorise l'ébauche d'un dictionnaire sémasiologique permettant de regrouper les différentes acceptions d'un terme défini (ex. pour *cabra*, outre le sens de « chèvre », on trouve, pour le moment, les sens de « chambrière du char », « chantier de tonneau», « chevalet pour scier », « faucheux (araignée) », déjà attestés dans le DOF, mais aussi « chouette », « criquet », « sauterelle » etc., non attestés.

4) enfin, un dernier « étage » s'ajoute à l'édifice, avec les étymons coiffant l'ensemble des lemmes issus d'une même source étymologique (ex. avec l'étymon CAPRA, on retrouvera en sus des formes citées plus haut : *cabraire*°, « chat-huant », *cabrarèla*°°, « chouette », *cabrèl*, « chevreau », *cabret*°°,

« billot », *cabreta*, « chèvre » et « chevreau », *cabri*, « chevreau », *cabridar*, « mettre bas (chèvre) », *cabrilha*, « chevreau », *cabrilhar*°°, « mettre bas (chèvre) », *cabrilhon*, « chevreau », *cabrilhonar*°°, « mettre bas (chèvre) », *chèvre (fr)*, « chevalet pour scier » et « chèvre », etc. (= champ étymon) (cf. plus loin pour la signification des ° et des °°).

2.3 L'implémentation des lemmes :

Le travail d'implémentation de la lemmatisation s'effectue à deux niveaux :

- le premier niveau se trouve en mode « opérateur de saisie » : la fiche d'implémentation comporte une rubrique « lemme » à liste déroulante permettant, soit de sélectionner une forme existante, puisée dans le DOF, ayant été intégrées lors de saisies précédentes, soit de saisir une nouvelle forme en suivant des règles sur lesquelles nous reviendrons plus loin, forme qui s'ajoutera automatiquement à la liste pour devenir disponible à son tour. Le remplissage de cette rubrique n'est bien sûr pas obligatoire pour la validation de l'enregistrement, cette absence de rattachement (résultat d'une ignorance ou d'un doute) a cependant pour conséquence de rendre incomplètes les cartes de synthèses lexicographiques basées sur la présence des lemmes.

- le second niveau, accessible seulement en mode « administrateur », permet par une fonction spéciale d'accéder à toutes les formes phonétiques différentes saisies pour une notion donnée. En regard, se trouvent l'ensemble des lemmes intégrés par les opérateurs. Il suffit alors de faire coïncider les formes phonétiques avec l'un des lemmes existants ou de créer un lemme pour que l'ensemble des formes phonétiques sous-jacentes similaires soit pourvu de la forme convenable. Cette fonction a pour avantage de donner une vision globale des variantes en présence, ce qui souvent est déterminant pour le choix de tel ou tel lemme, d'autre part elle a un caractère non définitif, puisqu'il est toujours possible de revenir sur telle ou telle forme.

2.4 Les règles

Cette tâche de lemmatisation pourrait presque s'automatiser en terme de recherche : forme phonétique → graphie phonologisante (dite par approximation « mistralienne ») → graphie classique → recherche automatisée dans la version informatisée du DOF = absent/présent.

Cependant, cela n'est pas aussi simple, car à la dimension présence/absence, s'ajoute la dimension de l'adéquation sémantique : en fait, pour lemmatiser, il faut se poser deux questions essentielles : d'une part, le terme existe-t-il ou non dans le DOF ? (ce qui suppose quand même de l'avoir reconnu en le « déshabillant » de sa variation dialectale et donc de bien connaître les parlers traités (cf. les exemples donnés pour « chèvre ») et d'autre part, s'il est attesté, a-t-il le même sens ?

Après la phase de reconnaissance de la forme phonétique (effacement de la variation), la lemmatisation proprement dite obéit aux principes suivants :

- les formes verbales sont ramenées à l'infinitif,
- les adjectifs et participes sont ramenés au masculin singulier,
- les substantifs au singulier (sauf si le singulier est inusité ou lorsque la forme au pluriel apporte un sens particulier).

Pour la confrontation avec le DOF, dans l'état actuel des choses, plusieurs options se présentent :

1) la forme reconnue est attestée dans le DOF avec le sens stipulé par la source (ici en l'occurrence le titre de la carte de l'atlas traitée), alors le terme figure tel qu'il est attesté (ex. *cabra* est attesté au sens de « chèvre », pour reprendre les exemples cités plus haut)

2) la forme reconnue est attestée dans le DOF, mais le sens spécifique contenu dans la base est différent, cette différence est marquée par la présence d'un symbole ° après le terme (ex. *cabra* est bien attesté aux sens de « chèvre », de « chevalet pour scier » et de « faucheux (araignée) », alors qu'il a été également recueilli avec les sens de « criquet », de « sauterelle » et de « chouette » dans les atlas, donc il figure

comme *cabra*° en regard des formes phonétiques recueillies avec ces sens spécifiques.

3) la forme reconnue n'est pas attestée dans le DOF, mais la base lexicale y figure avec d'autres affixes. On procède alors à l'ajout de l'affixe convenant, en respectant la graphie alibertine et l'on adjoint les symboles °° à la forme constituée (ex. *cabrilhar*°°, « mettre bas (chèvre) » n'est pas attesté, mais on y trouve *cabrilha* « jeune chèvre » et *cabrilhon* « jeune chevreau »).

4) la forme reconnue ne figure pas dans le DOF, hors homonyme. Cet ajout, en respectant la graphie alibertine est marqué par la présence du symbole * (ex. *reinard** ne figure pas comme entrée pour « renard » [21] ou encore *tondol** pour « assiette », etc.)

5) enfin les formes reconnues comme étant empruntées au français sont notées avec la graphie française et l'abréviation (fr) leur sont adjointes (ex. la forme phonique [ʃ'ɛv] est associée à *chèvre (fr)*).

Une sous-catégorie est apparue quand il s'est agi de traiter les formes relevées dans les atlas périphériques au regard de l'espace occitan (l'*Atlas du Centre* et l'*Atlas de l'Ouest* dont une partie du domaine comprend les parlers du Croissant qui ont été intégrés au réseau du Thesaurus Occitan). Il s'agit de termes inconnus du français standard, mais attestés dans les dictionnaires dialectaux. L'abréviation (fr*) leur est alors adjointe (ex. *encoubaisser (fr*)* « attacher une corne à une patte (pour entraver une bête) » attesté dans le *Glossaire du Centre de la France* de Jaubert).

Malgré tout, si la tâche semble facilitée par l'existence du dictionnaire de référence qu'est le DOF, il n'est pas toujours facile de se résoudre à une lemmatisation trop générale. Même L. Alibert n'a pas toujours tranché : soit il a utilisé la double vedette (ex. *lach ~ lait, mossilhar ~ morsilhar* « mordiller », etc), soit il a

[21] Cependant, le terme est attesté par *coa de reinard* « amarante » sous **coa** ce qui relève soit de l'oubli dans la part du travail du lexicographe, ou plus vraisemblablement du purisme (Alibert n'indique pas en général les emprunts récents au français).

fait usage des variantes (ex. *avelana* « noisette », Var. *aulana, auglana, avelan*) (voir plus loin).

Et comment ne pas évoquer les formes hybrides rencontrées dans les aires des atlas périphériques (Croissant et « amphizone » franco-provençale) où l'hybridité peut être phonétique comme dans le cas de [ʃjˈɛb] « chèvre » à lemmatiser sous *cabra* ou sous *chèvre (fr)* ou lexicale comme [asãbjˈad] *assemblada* « fête locale » qui est une forme occitane d'un terme largement répandue dans les parlers français de l'Ouest : l'*assemblée*.

2.5 Quelques statistiques dans le *Thesaurus occitan* :

Actuellement (avril 2013), le *Thesaurus occitan* comporte environ 1.212.250 formes phonétiques distinctes (techniquement des « enregistrements ») dont environ les deux-tiers (801.633) sont associés à un lemme, au nombre de 44.000, répondant au critère "forme/sens unique".

- 7429 lemmes (17%) apportent un changement de sens par rapport aux formes attestées dans le DOF,
- 6046 lemmes (13%) se différencient la forme attestée dans le DOF par la présence d'un affixe,
- 904 lemmes (2%) ne sont pas attestés dans le DOF.

Ce dernier chiffre est bien sûr très en-deçà de la réalité, puisque lorsque le lemme reconstitué (ou considéré comme tel) ne répond pas aux options 1, 2 ou 3, par réflexe ou par prudence, les opérateurs laissent la rubrique vide.

À l'inverse, la proportion de lemmes étiquetés « (fr) » (français) est très grande, 6589 (15%), puisque ces formes sont facilement reconnaissables et donc traitées en priorité ; la proportion de ce type devrait fortement baisser au fur et à mesure de l'implémentation des lemmes.

Enrichissements du DOF

Quelques exemples permettront de constater dans quelle mesure le *Thesaurus occitan* produit un enrichissement du DOF.

Enrichissement sémantique :

Sous la vedette **cuca**, *f,* du DOF, on relève les sens de 'lente', 'chrysalide', 'vermisseau', 'chenille', 'mite', 'artison' et 'petit insecte'. Idem dans le *Trésor du Félibrige* de Mistral, qui ajoute les sens de 'ver luisant' et de 'femme perfide' et fait des entrées à part (omises par L. Alibert) pour **cuco** 'rainette', ainsi que pour **cucho, cuco, cusso** 'tas, monceau, butte'.

Le Thesoc atteste bien sûr des sens relevés par Mistral et Alibert, « asticot », « chenille », « insecte en général », « lente », « ver luisant » et « grenouille « d'une part et « meule de paille », « petit tas de gerbes » d'autre part, mais atteste également de « courtillière », « hanneton », « ver blanc » et « crapaud », et en restant dans l'aire animalière délimitée par les lexicographes, des sens de « vipère » et d'« orvet », de « pansement au doigt », et d' »épouvantail » pour l'autre terme.

Autre exemple, sous la vedette **bramar**, *v intr,* du DOF, on relève les sens de « bramer; braire; mugir; beugler, s'égosiller ». Idem dans Le Trésor de Mistral qui ajoute « brailler, crier, rugir ; pleurer ; vociférer ».

Le Thesoc atteste bien sûr des sens relevés par Mistral et Alibert, mais élargit le sens à « hennir », « bêler », « chevroter », « aboyer », « croasser », « crier (de l'aigle ») et aussi « crier (charivari de mariage) », et « publier les bans ».

Enrichissement lexical par apport d'affixes

La vedette **pelha**, f, du DOF est déjà riche d'un grand nombre de dérivés : *pelhar, pelhaire, pelhandra, pelhandran, pelhandrós, pelhard, pelharòc, pelharocaire, pelhàs, pelheret, pelhièra, pelhotassa, pelhon, pelhasson, pelhomàs, pelhòc, pelhòl,-a, pelhòfa, pelhagondrit, pelhós, pelhòt, pelhum*. Il faudra en adjoindre d'autres avec les données du Thesoc : *pelharaud* « créature fabuleuse », *pelharòt* et *pelharotaire* « chiffonnier » (var. de *pelharòc*, *pelharocaire*), *pelharut* « déguenillé », *pelhatge* « linge », *pelhet* « chiffon ».

Enrichissement lexical :

Les absences lexicales du DOF, qui par ailleurs sont parfois attestées dans le *Trésor*, sont quelquefois des oublis comme *cantar* ou *reinard** (encore que ce dernier soit attesté par *coa de reinard*), mais le plus souvent sont des formes qui soit n'appartiennent pas au domaine couvert par le DOF, comme *pacha** « fesse », *masada** « fourmi »… soit des formes considérées comme des gallicismes et écartées pour cette raison dans un ouvrage dont la perspective est explicitement normative, comme par exemple *fermar* « fermer ». Alibert a en effet pour habitude de ne pas citer les gallicisme qu'il recommande d'éviter : sa liste de gallicismes corrigé ne donne que les correction et laisse les gallicisme implicite. Il est évident qu'un dictionnaire descriptif de l'occitan doit intégrer (et donc lemmatiser) les gallicismes : il faut des entrées *boèta, volur, achetar*… dans un dictionnaire général de la langue occitane.

Application : la lemmatisation d'*avelana*

La carte noisette de l'ALF est un bon exemple de carte complexe à la fois lexicalement et phonologiquement.

Un premier niveau de lemmatisation est donné par le passage à une forme orthographique classique : il livre des formes ***avelana***, mais aussi *auglana*, *aulana*, *averana*, *aurana*, *averaa*… Il livre aussi des formes du type *aulanha* ou *ametlana*. Il livre enfin *nosilha*, *noseta* et *anuçòla*. On ne trouve pas du côté du fruit de formes apparentées à *vaissa* ou à *còure* qui peuvent désigner l'arbre.

La graphie réduit sensiblement la dispersion des formes mais ne fournit pas d'emblée toute la structure lemmatique dont on peut avoir besoin.

Les premières formes citées ont en commun d'être des avatars de la même matière étymologique. Elles ne diffèrent que par l'intervention de processus phonologiques : traitement d'-LL- géminé intervocalique, chute de la prétonique interne, épenthèse consonantique, chute d'-*n*- intervocalique… Ces formes pourraient être lemmatisées par leur étymon, ABELLANA. On a vu

les difficultés éventuelles de cette approche. Elles peuvent être lemmatisées par une d'entre elles, préférablement une forme qui n'a pas subi les divers processus responsables de la divergence, soit *avelana*.

Dans un dictionnaire panoccitan complet, on pourrait admettre que la seule forme ***avelana*** soit suivie d'un article qui la définisse et l'illustre de citations. *Aulana, auglana, averaa* etc. seraient renvoyées à cette forme au titre de variantes (sauf spécificité sémantique liée exclusivement à une des formes qui justifierait un article propre en plus du renvoi).

Le renvoi au titre de variante est distinct du renvoi au titre de synonyme : c'est ce dernier qu'il faut envisager pour *noseta*, *nosilha* ou *anuçòla*. C'est aussi comme des synonymes (et non comme des variantes) qu'il faut considérer toute forme dont la composition morphologique ou l'étymon diffère. Les formes *avelanha, aulanha…* qui remontent à *ABELLANEA et non ABELLANA (à moins qu'elles ne résultent d'une fausse régression à partir d'*avelanièr* cf. plus loin), doivent être considérées comme des synonymes. De même *avelanilha* < °ABELLANICULA, *aulinha* < ABELLĪNEA, *averai* < ?°AVERATICU… De même *ametlana* issue d'un croisement (linguistique) d'*avelana* avec *ametla* (< AMYGDALA). De même encore les emprunts serviles du type *noaseta*.

On pourra éventuellement considérer que, en plus d'être lemme orthographique direct d'un bon nombre de variantes phoniques (`[aβel'ano], [avel'ano], [avel'ana]` …) et qu'en plus d'être le lemme lexical qui réunit des variantes marquées par des processus évolutifs lourds que la graphie enregistre, ***avelana*** peut aussi représenter la famille des mots issus d'une base °**ABEL-**[22] ? Comme *nosilha* peut être pris pour représentant des formes diminutives tirées de NUX.

[22] On explique classiquement ABELLANA et ses variantes comme des formes dérivées d'un toponyme: *abellana nux*, 'noix d'Abella' ville de Campanie réputée pour ses fruits (FEW 24.134, p. 28). Caton distingue ainsi "nuces abellanae, praenestinae, calvae…" (*De agricultura* 8.2, 133.6). On remarque toutefois que ABEL- est une désignation indo-européenne répandue du 'fruit' (germ. *apple, Apfel* 'pomme', breton : *aval* 'pomme', gaélique d'Écosse *ubhal*¸ russe *yabloko etc.*). Il est donc plus satisfaisant de voir dans *abellana nux* la désignation une noix éminemment comestible (noix-fruit) qui a rencontré un

Enfin, à un niveau ultime « avelana » avec des guillemets est un bon candidat pour traduire le français « noisette » et désigner en occitan le concept dont l'ALF ou les atlas régionaux relèvent la variété des désignations.

L'articulation plus précise des formes liées au lemme lexical *avelana* est donnée dans les tableaux 1 & 2. Ces tableaux sont classificatoires et non pas chronologiques. Ils montrent la cooccurrence dans une forme de divers processus évolutifs, non l'ordre dans lequel cette évolution s'est faite.

toponyme lui-même formé sur une base faisant référence aux fruits (cf. Gaffiot et l'explicitation virgilienne qu'il cite : *malifera Abella* 'Abella riche en fruit'). Une situtation initiale où *abelana nux* 'noix fruitière' est parallèle à *Abella* '(terre/pays) du fruit, abondant' fait naître une étymologie populaire où le nom du lieu explique le nom du fruit. De même gr. κάστανα et de-là lat. *castanea* est quelque fois expliqué par un toponyme pontique ou thessalien, mais il est plus convaincant de considérer que les toponymes viennent de l'arbre dont le nom, outre l'arménienr kask 'châtaigne' peut être rapproché du nom celtique du chêne *cassanos*, oc. *casse* (cf. Chantraine 1980 s.v. κάστανα, Harper s.v. *chestnut*). Un indice interne à l'occitan vient appuyer cette hypothèse: il existe un adjectif *abelan, -a* que Mistral applique à la terre et fait rimer au féminin avec *avelana* justement. Mistral traduit « la tèrra es abelana » par "la terre est généreuse". Le *-b-* d'*abelan* peut s'expliquer par un passage par une autre langue que le latin, ou un emprunt interdialectal.

TABLEAU 1 : ABELLANA (NUX) → *avelana* (et variantes)

chute d'n intervocalique n → Ø/ v_v	syncope e → Ø/	épenthèse vélaire Ø → g/w_l(r)	rhotacisme -LL- → r (/l)/v_v	palatalisation -LL- → ʎ	aphérèse a- → Ø/#_C	aboutissants (en notation orthographique)	autres évolutions [23]
+	-	-	+	-	-	*averaa*	*arevaa* (a)
+	-	-	-	-	+	*'velaa*	
+	+	-	+	-	-	*auraa*	
+	+	-	-	-	-	*aulaa*	
+	+	+	+	-	-	*augraa*	
+	+	+	-	-	-	*auglaa*	
-	+	+	-	-	-	*auglana*	*anglana* (b)
-	+	-	+	-	-	*aurana*	*auverana* (c)
-	+	-	-	-	-	*aulana*	
-	-	-	-	-	-	***avelana***	
-	-	-	-	-	+	*'velana*	
-	-	-	+	-	-	*averana*	*auverana* (c)
-	-	-	+	-	+	*'verana*	
-	-	-	-	+	-	*avelhana*	

[23] Autres évolutions : (a) metathèse : v(u)... l(r) → l(r)… v (u), (b) substitution de coda : n → w, (c) croisement de formes : *averana* × *aurana* → *auverana*, (d) croisement de formes : *aulanha* × *auranha* → *aurlanha*,

TABLEAU 2 : ABELLANEA (NUX) → *avelanha* (et variantes) (ou ***avelanièr*** interprété ***avelanhièr*** donne ***avelanha*** par fausse régression)

	syncope	épenthèse vélaire	rhotacisme		relèvement a→ i/ _ɲ		
	-	-	-		-	***avelanha***	
	+	-	-		-	*aulanha*	*aurlanha* (d), *alaunha* (a)
	+	-	-		+	*aulinha*	
	+	-	+		-	*auranha*	*aurlanha* (d), *araunha*
	+	+	-		-	*auglanha*	

Avertissement :

Les deux tableaux précédents classent les formes selon les processus phonologiques qu'elles présentent. La succession des colonnes ne cherche pas à refléter la chronologie relative des changements. En fait, cela ne peut pas être le cas parce que, par exemple, la syncope ne saurait précéder le rhotacisme qui suppose que la latérale géminée soit intervocalique. Plus généralement, le fait que la variation repose parfois sur un ordre différent des processus envisagé exclut que l'on puisse toujours ordonner de manière chronologique ainsi les processus pour un groupe de formes.

L'ordre des colonnes et des valeurs dans les colonnes ne tend qu'à mettre en évidence les regroupements qui apparaissent sur les cartes (Cartes 1 & 2 en annexe). Ces cartes distinguent tout d'abord les formes du type *avelana* des formes issues de dérivés de NUX : *nosilh, noseta, noisette*. Parmi les formes issues d'ABELLANA ou apparentées, on distingue d'abord par la couleurs celle qui supposent soit un étymon spécifique *ABELLANEA comme l'admet le FEW, soit une fausse régression à partir du nom de l'arbre (*avelanièr*, analysé *avelanhièr*, homophone et donnant *avelanha*). On distingue ensuite des formes gasconnes où la chute de l'*n* intervocalique donne des oxytons, souvent réinterprétés en masculins. L'ALF relève une seule forme féminine de ce type alors que l'ALG fait état d'une fréquent indécision sur le genre de la forme. La graphie classique de l'occitan notera de manière distincte et permettra donc de lemmatiser : *averaa* une forme féminine du type [awerãŋ] et *averan* une forme masculine de même réalisation.

Les aboutissants de ABELLANA présentent de manière attendue deux faits typiques du domaine gascon, le traitement de -LL- intervocalique latin (-LL- → -r- / v__v) et la chute d'*n* intervocalique (-n- → Ø/ v__v). On s'attend donc à trouver un type gascon *averaa* [awerˈãː] en face d'un type général *avelana* [aβelˈanɔ] etc. Éventuellement, on s'attend à ce que les deux traits gascons puissent ne pas coïncider toujours. De fait, on trouve *averana* avec -LL- → -r- mais sans chute d'*n* intervocalique, mais on ne trouve pas d'**avelaa* suggérant que

l'aire de la chute d'*n* intervocalique est incluse dans celle du traitement gascon d'-LL- géminé.

Un fait massif vient compliquer ce tableau, la syncope de la prétonique interne dans nombre de parlers (*aulana*), suivie de l'éventuelle épenthèse d'une vélaire (*auglana*). La chute de la prétonique interne est de règle en occitan, en syllabe ouverte. La question est précisément celle-ci : la seconde syllabe d'ABELLANA est-elle ouverte ou fermée ? Tant qu'on a affaire à une géminée réelle, elle est fermée, mais elle peut être ouverte à partir du moment où l'opposition -LL- géminé -L- simple se transforme en opposition qualitative (*l* alvéolaire ou rétroflexe contre *l* vélaire par exemple…). On admettra donc que la syncope s'est produite dans des parlers qui ont réduit précocement la géminée au profit d'une différenciation qualitative ou pour aboutir à une neutralisation.

Le fait qu'on trouve des formes du type *aulaa* en gascon occidental peut s'expliquer si l'on pose une syncope précoce dans ces parlers, au stade où l'opposition -LL- / -L- avait pris la forme -l- / −ɬ−. *l* aurait ainsi été gelé à ce stade (puis vocalisé) au lieu de suivre l'évolution générale vers r. *auraa* suppose une syncope soit au stade cacuminal soit au stade rhotique : ABELLANA → aβelˈana → awlˈana → awrˈana ou ABELLANA → aβelˈana → aβerˈana → awrˈana.

Ailleurs qu'en Gascogne, l'opposition du type *avelana* au type *auglana* peut être imputé aussi à une chronologie relative différente de ce qui est ici une perte non compensée de la géminée intervocalique et de la syncope. On peut risquer que les parlers qui en finale traitent de même -L- simple et -LL- géminé (ou *gal* rime avec *ostal*) sont de bons candidats à une neutralisation précoce, ceux qui distinguent le produits d'L simple et LL géminé (qui disent *ostau* mais *gal*) sont, au contraire, candidats à une plus longue préservation de la géminée. Le languedocien central traite de manière identique -L- et -LL- en toute position. Le languedocien oriental et le provençal nîmois présentent un traitement différencié. On peut donc avancer l'hypothèse que dans cette zone *abellana* a conservé une géminée assez longtemps pour résister à la syncope de la prétonique interne.

Les aires d'*aulana*, *auglana* d'un côté, *avelana* de l'autre, ne répondent pas trop mal à cette explication proposée sauf que l'aire d'*avelana* s'étend assez largement en zone de confusion d'-L- et de -LL-. Outre que la distinction en finale n'est qu'un indice et non une preuve et donc qu'elle n'est pas non plus une contrainte sur le déroulement chronologique, on peut supposer aussi des rediffusions de dialecte à dialecte. On peut le supposer, d'une part à la vue de la forme de l'aire *abelana* qui avance dans les vallées fluviales. On le peut aussi en s'appuyant sur une remarque de l'auteur anonyme d'un dictionnaire occitan français manuscrit de la toute fin du 18e siècle (Bazalgues 1987). Ce dictionnaire a été composé à Saint-Hippolyte-du-Fort, dans le piémont cévenol. L'auteur dans son article *avelano* explique la différence qu'il y a en français entre « noisetier » cultivé et « coudrier » sauvage. Il en tire la conclusion que sur la montagne de la Fage (qui domine Saint-Hippolyte-du-Fort) il n'y a que des « coudrettes », des bois de coudriers, et pas de noisetiers. Or ce pays haut où croissent des « coudriers » est un pays où on dit *auglana* et non plus *avelana* comme à Saint-Hippolyte. On peut donc imaginer que les *avelanas* avancent là où l'arbre est cultivé ou inconnu. Dans le dernier cas la noisette vendue comme fruit sec a pu être le vecteur du mot : Jean Michel au vers 2276 de son *Embarràs de la fièira de Beucaire* évoque des « vendeires d'avelanas », attestant qu'on vendait des noisettes à la foire de Beaucaire (cf., commodément, Gardy 1974). Et une expression comme *cracar averaas* en Béarn, c'est à dire *crocar d'avelanas* au sens du français *boire du petit lait* (cf. Lespy & Raymond 1887 s.v. *aberaa*) suggère que la noisette a pu être une friandise très prisée et donc commercialisée, et que ce commerce a pu faire voyager certaines formes (comme *avelana*).

Conclusion

Toutes les langues qui ont une orthographe d'usage établie trouvent dans les formes de cet usage un outil commode pour trier et référencer la masse de leurs variantes dialectale. Les

linguistes qui pratiquent ces langues sont aussi des usagers (et habituellement des usagers experts) des codes graphiques communs. En domaine occitan cette compétence d'une part n'est pas systématique (il y a des linguistes travaillant dans le domaine occitan qui ont une connaissance très approximative des graphies d'usage) et d'autre part est sous exploitée par ceux qui la maîtrisent par crainte de sortir d'une supposée réserve ou neutralité scientifique que garantirait le non recours aux graphies d'usages de la langue. La présente contribution a voulu souligner l'intérêt d'une *normalisation* de la linguistique occitane, au sens de l'entrée sans confusion mais sans réserve dans une pratique *normale* dans d'autres domaines linguistiques, où code orthographique et notations phonétiques se complètent. Quelle que soit la graphie de l'occitan, la démarche serait soutenable. Elle l'est pour l'anglais ou le français dont on connaît les orthographes complexes. Elle le serait avec la norme mistralienne (comme l'utilise Ronjat dans sa grammaire (h)istorique, qui illustre d'abord ses analyses avec les formes rhodaniennes, malgré leur non-centralité dans le diasystème occitan, le provençal rhodanien étant connu pour avoir développé un bon nombre d'innovations). Toute standardisation en place, quelques réserves techniques que l'on puisse faire à son sujet, pourrait être opérationnelle. Il se trouve toutefois que la graphie dite classique (prolongée par une normalisation linguistique pluricentrique) est aussi un instrument techniquement assez commode et pertinent de manipulation et d'analyse des données comme on a tenté de le montrer à propos des variantes d'*avelana*. C'est aussi un instrument que l'entreprise du THESOC met à son service dans son projet global.

Bibliographie :

Alibert, L. (1966). *Dictionnaire occitan-français, sur la base des parlers languedociens.* (DOF) Toulouse : Institut d'Estudis Occitans.

Alibèrt, L. (1976). *Gramatica occitana, segon los parlars lengadocians.* Montpelhièr : CEO. [reedicion corregida: 1ra edicion 1935].

Bazalgues, G. éd. (1974). *Dictionnaire Languedocien-Français manuscrit* (Saint-Hippolyte-du-Fort 1798) [anonyme] Montpellier : CEO-UPV.

Boisgontier, J. (1981-1986). *Atlas Linguistique et Ethnographique du Languedoc oriental*. (ALLOr) Paris : CNRS, 3 vol.

Brun-Trigaud, G. (2003a). « Présentation du logiciel d'indexation des atlas linguistiques régionaux ». *VIIe colloque international de dialectologie et de littérature du domaine d'oïl occidental : A l'ouest d'oïl, des mots et des choses* ... , édité par S. LAINE, P. BOISSEL et C. BOUGY, 293-299. Caen : Presses universitaires de Caen.

Brun-Trigaud, G. et F. CARTON (2003b). « Lemmes, supra-lemmes : dilemmes... Problèmes d'indexation de l'Atlas linguistique picard et de l'Atlas linguistique du Centre », *Sempre los camps auràn segadas resurgantas. Mélanges offerts à Xavier Ravier*, édité par Jean-Claude BOUVIER, Jacques GOURC et François PIC, 63-72. Toulouse : CNRS - Université Toulouse-Le Mirail.

Carton F. et A. Dawson (2010). *Index lemmatisé et étymologique de l'Atlas Linguistique et ethnographique Picard*. Amiens : Université de Picardie Jules Verne.

Chantraine, Pierre 1970-1980 *Dictionnaire étymologique de la langue grecque : histoire des mots*, Paris : Klincksieck, 2 vol. xviii, 1368 p.

Dalbera, J.-Ph. (1998). « La base de données Thesoc : état des travaux » in J. Gourc et F.Pic (éds) *Toulouse à la croisée des cultures* : 403-417.

Dauzat, A. (1929). « Essais de géographie linguistique (nouvelle série) » *Revue des langues romanes* 66 : 45-80.

Dondaine C. (2002). *Trésor étymologique des mots de la Franche-Comté, d'après l'ALFC*. Strasbourg : Société de linguistique romane.

Dubuisson P. (1976-1982). *Atlas Linguistique et Ethnographique du Centre*. Paris : CNRS (3 vol.).

Gardette P. (1976). *Atlas Linguistique du Lyonnais. V. Commentaires et Index*. Paris : CNRS.

Gardy, Ph. ed. (1974). *L'embarràs de la fièira de Beucaire de Jean Michel*. Montpelhier : Centre d'Estudis Occitans.

Gilliéron J. et E. Edmond (1912). *Table de l'Atlas Linguistique de la France*. Paris : Champion.

Gillieron, J. et E. Edmont (1902-1910). *Atlas linguistique de la France*. (ALF) Paris : Champion.

Harper, Douglas 2003 *Online Etymology Dictionary* (http://www.etymonline.com : consulté le 2013-07-22), Ohio University.

Jaubert H.-F. (1864-1869). *Glossaire du Centre de la France*. Genève : Slatkine Reprint (1970).

Kremnitz, G. (1974). *Versuche zur Kodifizierung des Okzitanischen seit dem 19. Jahrhundert und ihre Annahme durch die Sprecher*. Tübingen : G. Narr.

Lafont, R. (1971). *L'ortografia occitana, sos principis.* Montpelhièr : CEO.

Lafont, R. (1972). *L'ortografia occitana, lo provençau.* Montpelhièr : CEO.

Lafont, Robert (1997). *Quarante ans de sociolinguistique à la périphérie*. Paris : Harmattan.

Lespy, V. & P. Raymond (1887). *Dictionnaire Béarnais ancien et moderne*. Montpellier : Hamelin.

Massignon G. et B. Horiot (1971-1983), *Atlas Linguistique et Ethnographique de l'Ouest*. Paris : CNRS (3 vol.).

Mistral, F. (1882-1886). *Lou tresor dóu Felibrige*. (2 vol.) Aix-en-Provence : V^{e} Remondet-Aubin. [réimp. avec une préface de J.Cl. Bouvier 1979, Aix : Edisud.]

Palay, S. (1932-33). *Dictionnaire du béarnais et du gascon modernes*, Pau : Marrimpouey. [réédition Paris : CNRS 1963, 1994)

Ravier, X. *et alii* (1978-1993). *Atlas linguistique et ethnographique du Languedoc occidental.* (ALLOc) Paris : CNRS.

Ronjat, J. (1930-41). *Grammaire istorique (sic) des parlers provençaux modernes*, Montpellier : Société des Langues Romanes.

Sauzet, P. (2002). « Réflexions sur la normalisation linguistique de l'occitan » in D. Caubet, S. Chaker & J. Sibille éds *Codification des Langues de France.* Paris : L'Harmattan, 39-61.

Schlieben-Lange, B. (1971). *Okzitanisch und Katalanisch : ein Beitrag zur Soziolinguistik zweier romanischer Sprachen*. Tübingen : G. Narr.

Séguy, J. (1954-1973) . *Atlas linguistique et ethnographique de la Gascogne*. Paris : CNRS.

Sumien, D. (2006). *La standardisation polycentrique de l'occitan. Nouvel enjeu sociolinguistique, développement du lexique et de la morphologie*. Turnhout :Brepols.

Taupiac, J. (1980). « Quin modèl lexicografic foguèt Simon Jude Onorat per Frederic Mistral. » *Quasèrns de Linguistica Occitana* 9 : 17-22.

Taverdet G. (1988). *Index de l'Atlas Linguistique de la Bourgogne*. Dijon : ABDO

Taverdet G. (1989). *Index de l'Atlas Linguistique de la Champagne et de la Brie de Henri Bourcelot*. Dijon : ABDO.

Taverdet G. et P. Dubuisson (1993). *Index de l'Atlas Linguistique du Centre*. Dijon : ABDO.

Teulat, R. (1980). « Remarcas subre l'ortografia del *Tresor dóu Felibrige*. » *Quasèrns de Linguistica Occitana* 9 : 52-57.

THESOC http://thesaurus.unice.fr/

ANNEXES

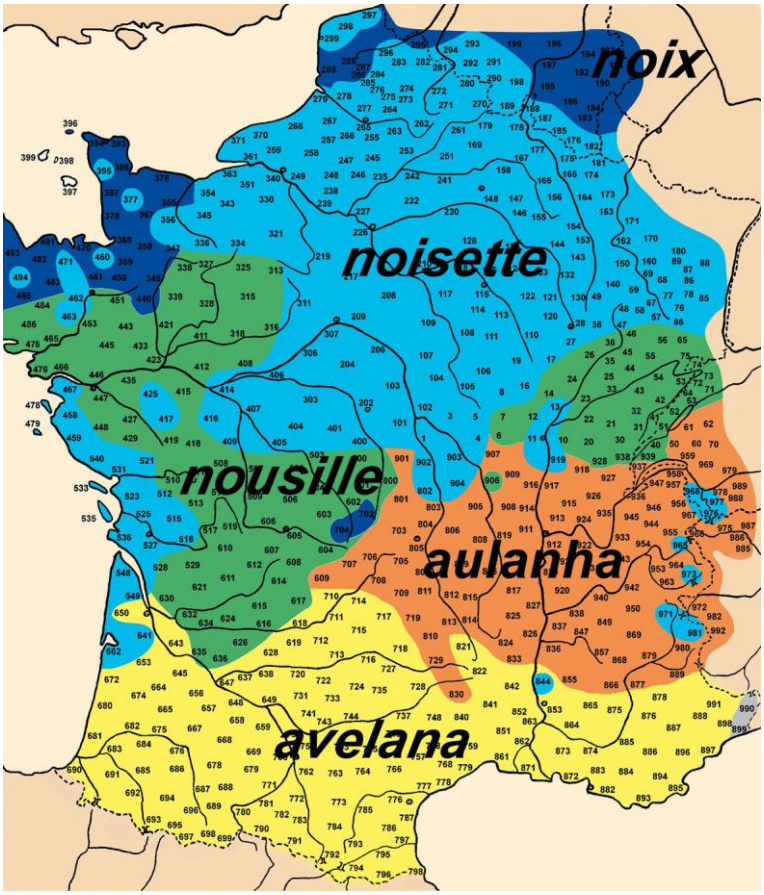
TABLEAU 3

aboutissants (en g.c.)	autres evolutions
→ **avelana**	
→ ******avelaa*	
→ averana	→ auverana[24]
→ averaa	→ arevaa[25]
→ aulana	
→ auglana	→ anglana[26]
→ aulaa	
→ auglaa	
→ aurana	→ auverana[27]
→ ******augrana*	
→ auraa	
→ augraa	
→ ‘velana	
→ ‘velaa	
→ ‘verana	
→ *****‘*veraa*	
→ avelhana	

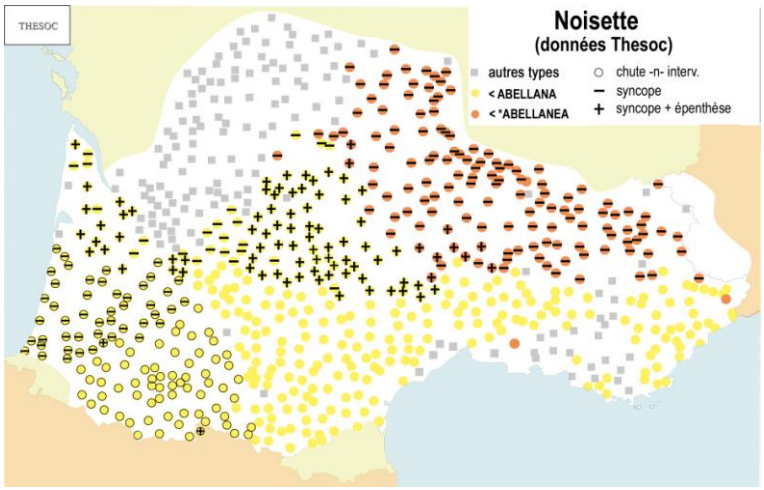
[24] < *averana* × *aurana*

[25] metathèse

[26] substitution de coda : w → n



Carte 1 : Dénominations de la noisette (ALF 919)



Carte 2 : Extension du lemme « avelana » selon les données du Thesaurus Occitan