

Histoire du corpus d'Amsterdam: Le Traitement des données dialectales

Yves Charles Morin

1 Le projet de Dees: une dialectologie de l'ancien français

L'objectif fondamental de Dees était de faire une dialectologie rigoureuse de l'ancien français en utilisant des échantillons représentatifs de toutes les sources cohérentes (cf. Dees 2001). Il avait fait porter son action sur les trois suivantes:

- les chartes régionales originales, datées et localisées,
- les manuscrits dont les graphies sont suffisamment homogènes pour permettre de localiser le copiste (ou les copistes de même profil) ayant laissé un témoignage cohérent d'un usage régional spécifique,
- les œuvres poétiques, pour leur métrique et leurs rimes.

L'essentiel de son travail a été consacré aux deux premières. Les chartes datées et localisées constituent les « témoignages primaires » nécessaires à l'interprétation dialectologique des secondes, qu'il qualifiait de « témoignages dérivés », et qui comprennent en particulier tous les manuscrits des textes littéraires. Il s'était aussi considérablement investi dans la stémattologie, en partie pour contrôler l'homogénéité des manuscrits des textes examinés.

Il avait fait sien la démarche de Fallot (1839), en l'élargissant et en la systématisant. Il considérait essentiel de réexaminer tous les résultats antérieurs de la recherche en dialectologie française et l'ensemble des critères que des générations de philologues avaient fini par transmettre. La dialectologie de l'ancien français, à ses origines, impliquait trois volets de recherche. Le premier s'intéressait à l'identification des caractéristiques graphiques des documents anciens dont on peut établir avec suffisamment d'assurance l'origine géographique et la date de rédaction — ces critères excluent la plupart des manuscrits des œuvres littéraires et une proportion importante des documents légaux, même si ce sont ces derniers qui en fin de compte constituent l'immense majorité des documents retenus. Le second était la reconstruction historique des prononciations des dialectes de l'ancien français à partir des dialectes modernes. Les premiers chercheurs se fondaient souvent sur leurs propres enquêtes des patois, alors bien vivants sur tout le domaine gallo-roman; enquêtes qui pouvaient alors porter précisément sur des aspects spécifiques de la prononciation dont ils avaient besoin pour l'interprétation des graphies des chartes. (La dialectologie historique moderne doit se satisfaire d'enquêtes qui n'ont pas été conçues spécifiquement pour répondre à ces préoccupations.) On notera à cet égard les travaux remarquables pour leur époque des Bonnier (1889–1890), Philipon (1910, 1912, 1914), Wailly (1870, 1871, 1876), ou Wilmette (1888). Enfin, le dernier volet établissait des correspondances entre les graphies anciennes analysées dans le premier volet et les prononciations reconstruites dans le second, permettant de faire de nouvelles hypothèses sur les prononciations anciennes qu'il aurait été difficile de reconstruire autrement. Il est important de souligner cependant que la nature des informations que l'on peut extraire des graphies pour la reconstruction des prononciations anciennes est problématique et qu'un débat a opposé les conceptions de Dees à celles de la scriptologie.

2 Les témoignages dérivés

Les manuscrits des œuvres littéraires sont le plus souvent exclus de la première étape de l'analyse dialectologique car, sauf exceptions, ils sont rarement datés et n'offrent aucune garantie d'être représentatifs d'un usage régional spécifique. La conformité d'un manuscrit avec un ensemble de graphies caractéristiques des chartes d'une région particulière à un moment donné, cependant, permet de faire l'hypothèse que les graphies de ce manuscrit, dans leur ensemble, sont représentatives de la pratique scripturaire de cette région à cette époque, ce qui permet de considérer ce document comme un témoignage « secondaire » ou « dérivé » de cet usage régional. On pourra utiliser de tels documents pour identifier de nouvelles caractéristiques graphiques régionales, non documentées ou mal documentées dans les témoignages primaires — les chartes, en effet ne se démarquent, ni par la variété de leur vocabulaire, ni par la richesse de leur style.

2.1 Le protocole de localisation

Le corpus d'Amsterdam des textes littéraires comprend un grand nombre de documents qui n'ont pas été retenus pour la confection de l'*Atlas des textes littéraires* (Dees 1987). Le corpus analysé pour cet atlas comprend 201 manuscrits dont les coefficients de conformité graphique avec les chartes d'une région spécifique sont supérieurs ou égaux à 70.

L'équipe d'Amsterdam avait développé plusieurs outils pour établir des degrés de conformité. Chacune des cartes publiées de l'*Atlas des chartes* (Dees 1980) peut être considérée comme une fonction ayant pour domaine 28 régions dont la valeur prise pour chacune d'elles est le pourcentage d'utilisation d'une certaine graphie. Par exemple, la carte 89 est la fonction: « \ulcorner ou \urcorner ~ \ulcorner o \urcorner de l'indéfini 'tout, tous, toute(s)' », qu'on peut lire ainsi: « proportion des graphies utilisant un digraphe du type \ulcorner ou \urcorner parmi l'ensemble de celles qui utilisent soit ce digraphe, soit la lettre simple \ulcorner o \urcorner pour l'indéfini 'tout, tous, toute(s)' ». Les types graphiques apparaissant ici entre les délimiteurs \ulcorner ... \urcorner sont des abréviations pour un ensemble de graphies jugées équivalentes. Celles-ci sont précisées sur chacune des cartes de l'*Atlas des chartes* (avec une liste complète des formes dans lesquelles elles apparaissent dans l'inventaire des pages 322 à 369). Dans cet exemple \ulcorner ou \urcorner note exactement le digraphe <ou> qu'on trouve dans les formes *thous, thoutes, tou, toul, tour*, etc. et \ulcorner o \urcorner la lettre <o> non suivie de <u, l, i, y> dans les formes *thos, to, tos, tot, tote*, etc.; le type \ulcorner aus \urcorner , pour la préposition contractée 'aux = à+les' que nous examinerons plus tard, est une abréviation pour l'ensemble des formes *aus, aux, auz* et *ax*.

Les cartes retenues dans l'*Atlas des chartes* sont limitées aux fonctions discriminantes: ont été exclus tous les usages graphiques dont les distributions sont relativement uniformes, ou qui n'apparaissent pas dans un nombre suffisant de chartes pour permettre des résultats significatifs. Une carte portant sur la distribution des graphies des voyelles dans les formes monosyllabiques du nom 'comte', par exemple, a peu d'intérêt pour la localisation, car ces formes sont pratiquement toutes *cuens* ou *quens*, et n'ont que très rarement des formes sans marque graphique de diphtongaison, comme *cons* ou *cuns*. (Cette uniformité, cependant, n'est pas indifférente pour l'histoire du français, quand on la compare à la variabilité bien affichée de 'bon' de la carte AC 120.) Nous appellerons ici « traits (graphiques) discriminants », ceux pour lesquels il a été possible de construire une carte suffisamment discriminante incluse dans l'*Atlas des chartes*.

Pour localiser un manuscrit, on compare les valeurs de ses traits discriminants aux valeurs moyennes des traits discriminants de chacune des régions du maillage géographique adopté. La région pour laquelle l'écart est le plus faible est la région d'origine la plus probable du manuscrit. Une mesure de conformité simple pourrait être le coefficient de corrélation décrit par Hubert et Karin van Reenen (1988) qui fait appel aux traits discriminants fournis par les 268 premières cartes (en ignorant les suivantes, qui portent sur les propriétés syntaxiques).

L'algorithme effectivement utilisé pour sélectionner les manuscrits de l'*Atlas des textes littéraires* est décrit en partie dans l'introduction de cet Atlas (Dees 1987: xxiv-xxvii) et en partie dans un article antérieur (Dees 1985). Il utilise seulement 261 traits discriminants (les sept autres n'ayant pas été suffisamment stables pendant le XIII^e siècle, comme il sera précisé plus bas), et s'appuie sur un maillage géographique de 87 points, beaucoup plus serré que celui des 28 régions documentées dans l'*Atlas des chartes*, et de plus incluant l'Angleterre, qui n'avait pas été examinée dans ce même atlas. Ce maillage est composé de toutes les régions et sous-régions pour lesquelles il existe au moins trente chartes dans le corpus. Le protocole de détermination du degré de conformité est relativement plus complexe que le coefficient de corrélation précédent, car il faut compenser l'indétermination (ou la détermination insuffisante) de certains traits discriminants, non seulement dans les manuscrits à localiser (ce qui arrive surtout pour les documents relativement brefs), mais aussi dans les chartes de certains points du maillage. (L'augmentation du nombre de points allant de pair avec la diminution des chartes associées à chacun d'entre eux.) Une correction statistique s'impose alors, qui fait appel à une moyenne pondérée des formes recueillies dans les points contigus.

Il ne fait aucun doute que la démarche entreprise par l'équipe d'Amsterdam représente une avancée majeure pour la fiabilité des décisions, dont toutes les recherches devront s'inspirer. La documentation maintenant lacunaire de plusieurs aspects du protocole de localisation — en particulier la pondération pour les données insuffisantes — et la perte de fichiers informatiques consignnant les résultats des analyses des 87 points du maillage rend néanmoins difficile le contrôle des résultats publiés et son application à de nouveaux manuscrits.

Une autre perte majeure est celles des fichiers relatifs à la lemmatisation du corpus des textes littéraires. Dees (1987) fait état de cette lemmatisation, que plusieurs chercheurs pouvaient encore utiliser à la fin des années 1980.

Nous avons pu en effet — et cela de façon très commode — passer en revue plusieurs millions de formes enregistrées dans la mémoire de l'ordinateur, chacune de ces formes étant rapportée soit à un lemme donné ('femme', 'beau', 'bien', etc.), soit à une classe morphologique (parfaits en *-erent*, adverbess en *-ment*, substantifs féminins en *-or*, substantifs masculins du cas sujet singulier). Chaque fois que, à l'intérieur d'un lemme (ou d'un ensemble de lemmes) ou à l'intérieur d'une catégorie morphologique, nous avons cru remarquer... ». Dees (1987: xviii)

2.2 Problèmes mineurs

Ces pertes sont fâcheuses mais contingentes. Le remède est relativement simple: reprendre les longues et laborieuses analyses dont on a perdu les traces. Avant d'examiner d'autres problèmes inhérents à la nature des données à analyser, il est bon de mentionner ici trois problèmes mineurs.

2.2.1 Informations contradictoires

Le premier concerne des informations contradictoires trouvées pour deux des manuscrits dans les fichiers informatiques. Le manuscrit du *Lancelot*, roman en prose du XIII^e siècle, ms. A, (éd. A. Micha, Tome VII, TLF, Paris/Genève 1978) reçoit une double localisation: (1) Somme et Pas-de-Calais, avec un coefficient de conformité de 76 (cf. Dees 1987: 521) et (2) Ardennes Sud (code régional 50) avec un coefficient de conformité de 63. De la même manière, le manuscrit *La vie de Sainte Marguerite de Wace*, ms. M (éd. E.A. Francis, CFMA, Paris 1932) a une double localisation: (1) Langres et environ, avec un coefficient de conformité de 84 (cf. Dees 1987: 529) et (2) Chaumont et environ (code régional 63), avec un coefficient de conformité de 67. Il peut y avoir eu un croisement entre plusieurs hypothèses de localisation. On doit très certainement accorder plus de valeur aux résultats publiés.

2.2.2 Reconstruction des manuscrits édités

Le second concerne la « reconstruction » des manuscrits édités. La localisation se fondant sur les caractéristiques de la graphie d'un manuscrit, il est essentiel de respecter celle-ci et de ne pas introduire ce nouvel intermédiaire qu'est l'éditeur moderne. Les leçons « fautives » rejetées ont donc été systématiquement restituées dans le texte à analyser. La reconstruction du manuscrit se fait à partir des notes données par les éditeurs eux-mêmes, et sa justesse dépend de leur exactitude, ce qui peut être problématique. Il n'est pas rare, en particulier pour les éditions qui se fondent sur un grand nombre de manuscrits, que les éditeurs n'indiquent pas les leçons rejetées lorsqu'ils considèrent qu'elles portent sur des détails graphiques secondaires — qui sont cependant essentiels pour la localisation. Ainsi, l'édition de Constans du *Roman de Troie* de Sainte-Maure ne permet pas de reconstruire de façon suffisamment précise la forme du manuscrit D55 de la bibliothèque Ambrosienne, comme on peut voir en comparant les seize premiers vers de la version « rétablie » par Dees à partir des indications de Constans, à ceux de l'édition partielle, plus récente, de Baumgartner (1998, *Livre de poche*):

715	peleus iert uns riches reis mout proz mout sages mout corteis par grece alot sa seignorie e del regne ot mout grant partie sa terre teneit quitement	Peleüs iert uns riches reis Mout prouz mout saives mout corteis. Par Grece alot sa seignorie E del regne ot mout grant partie. Sa terre teneit quitement,
720	bien e en pais e saivement icist reis aveit un son frere fiz de son pere e de sa mere eson iert par non apelez	Bien e en paiz e saivement. Icist reis aveit un son frere, Fiz de son pere e de sa mere: Eson iert par non apelez.
724	en penelope la citez ne sai s iert reis o cuens o du s quar le livres ne m en dit plus icist aveit un fiz eson	En Penelope la citez Ne sai s'iert reis o cuens o dux , Quar li livres ne m'en dit plus. Icist aveit un fiz Eson
728	qui esteit apelez jason de grant beaute e de grant pris e de grant sens si com jo truis	Qui esteit apelez Jason, De grant biauté e de grant pris E de grant sens, si cum je truis.

Le manuscrit « restitué » a été localisé en Vendée avec un coefficient de 61, trop faible pour le prendre en compte dans l'*Atlas des textes littéraires*. Une meilleure transcription aurait probablement permis de le localiser avec plus de précision.

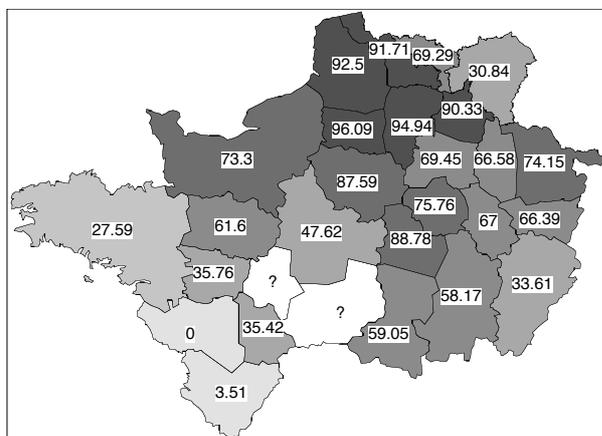
2.2.3 Résolution des abréviations

Les abréviations originales sont généralement résolues dans les éditions modernes, tant celles des chartes que des textes littéraires. Les abréviations ont toutes été remplacées dans les fichiers du corpus d'Amsterdam par leur « résolution » — c'est-à-dire par la graphie ordinaire « équivalente » du manuscrit —, même dans les rares cas où les éditeurs les avaient conservées, comme dans l'édition semi-diplomatique du *pseudo-Turpin* de Auracher (1877).

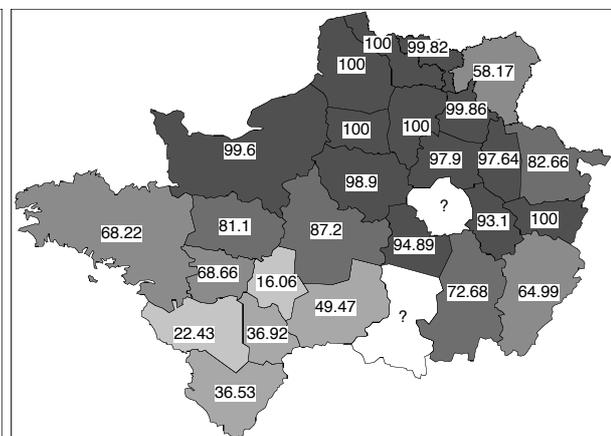
Ceci introduit une distorsion statistique, puisque c'est la forme non abrégée la plus fréquente qui est utilisée à la place de *toutes* les abréviations, distorsion qui semble le plus souvent mineure. Certains choix des éditeurs, cependant, peuvent être une source d'indétermination. Ainsi Andolf, dans son édition du ms. de *Floovant* (1941) a rendu le signe abrégé 7 par *et* quand il vaut 'et' et par *a* quand il vaut 'à' (cf. p. vi, cxxi). Les copistes du manuscrit (ils sont plusieurs) l'écrivent effectivement le plus souvent *a* et, exceptionnellement, *e* (2 occurrences) ou 7 (12 occurrences). L'éditeur estime que la prononciation de cette préposition devait déjà être [e] ou [ɛ], comme on l'entendait dans les patois bourguignons du début du XX^e siècle (ALF 525, 549, etc.), mais il n'a pas jugé approprié de résoudre 7 en *e* pour la préposition, comme il l'a fait pour la conjonction. Les fichiers du corpus d'Amsterdam se conforment systématiquement aux décisions prises par les éditeurs. Ce genre de décision explique peut-être pourquoi l'*Atlas des Chartes* examine les variantes graphiques «es» de la préposition contractée 'aux = à+les', mais n'a aucune carte pour la variante graphique «e» de la préposition simple 'à'.

2.3 Tenir compte de l'évolution

L'évolution des usages graphiques pendant la période considérée constitue un véritable problème pour toute entreprise de localisation des manuscrits. Certaines des cartes de l'*Atlas de chartes* font apparaître d'importants changements dans le temps de certains usages graphiques (cartes 34a/b, 37a/b, 87a/b, 89a/b, 101a/b, 162a/b, 266a/b). On a reproduit ci-dessous les cartes AC 89a et AC 89b: « ou^{r} ~ o^{r} de l'indéfini 'tout, tous, toute(s)».



Carte 1a. « ou^{r} ~ o^{r} de l'indéfini 'tout, tous, toute(s)'
1201-1275 – AC 89a



Carte 1b. « ou^{r} ~ o^{r} de l'indéfini 'tout, tous, toute(s)'
1276-1300 – AC 89b

Des caractéristiques dialectales qui évoluent rapidement dans le temps sont difficilement utilisables pour la localisation des manuscrits de textes littéraires, parce qu'on ignore le plus souvent non seulement le lieu où ils ont été copiés mais aussi la date à laquelle ils l'ont été. Ces caractéristiques dialectales instables dans le temps n'ont tout simplement pas été retenues pour la localisation (communication de Piet van Reenen au cours de cet atelier). On ignore cependant dans quelle mesure les autres traits discriminants ont été suffisamment stables pendant le XIII^e siècle. On peut être assuré que les chercheurs de l'équipe ont examiné tous les cas potentiels de variation dans le temps et qu'ils ont consigné dans l'*Atlas des chartes* tous les cas avérés de variation. On peut ainsi vérifier que le trait discriminant «¹cestes, celes¹ ~¹ces¹ du déterminant démonstratif féminin pluriel » (AC 75) est bien uniforme et qu'il n'y a pas de différences significatives entre les deux périodes 1201–1275 et 1276–1300 (sauf peut-être en Normandie, où la proportion des formes du type ¹cestes, celes¹ décroît de 21% à 1%).

Un nombre peut-être significatif de traits discriminants ont cependant pu être retenus, qui n'offrent aucune garantie absolue d'avoir été stables: ce sont ceux qui apparaissent dans un nombre insuffisant des chartes antérieures à 1275 pour qu'une évolution puisse être notée.

Idéalement, il faudrait élargir le corpus pour y inclure plus de chartes anciennes, s'il en existe. Les gains seraient malgré tout insuffisants, semble-t-il. Ainsi, pour la Bourgogne, le corpus d'Amsterdam comprend 79 chartes, dont seulement 9 pour la période 1201–1250. Philipon (1910, 1912), pour son analyse de l'ancienne langue en Bourgogne, a examiné 315 chartes du XIII^e siècle, dont 277 n'ont jamais été éditées et ne sont pas incluses dans le corpus d'Amsterdam (certaines n'auraient peut-être pas été admissibles); aucune d'entre elles cependant n'a été écrite avant 1251. Le tableau suivant donne une idée de la répartition dans le temps des chartes du corpus d'Amsterdam:

Période	1201–1225	1226–1250	1251–1275	1276–1300	total
Nb. de chartes	109	224	1514	1453	3300

On voit à quel point le nombre est limité des chartes antérieures à 1251, les seules qui permettraient d'isoler des caractéristiques dialectales fiables pour la localisation des manuscrits antérieurs à cette date. Pour la période 1201–1225, il n'y a que cinq régions parmi les vingt-huit qui rassemblent au moins six chartes (c'est le nombre minimum de chartes que Dees estimait nécessaire pour que des résultats soient significatifs): Charente et Charente-Maritime, Aisne, Nord, Hainaut, Moselle et Meurthe-et-Moselle. Cette pauvreté relative des témoignages primaires anciens explique peut-être pourquoi un nombre important des manuscrits les plus anciens n'ont pu être localisés avec un coefficient de conformité suffisamment élevé.

2.4 Textes contaminés

La « contamination des textes » est un autre problème majeur pour la localisation des manuscrits. Un texte est « contaminé » lorsqu'un copiste ou plusieurs copistes successifs n'ont pas été systématiques et que le manuscrit dont on dispose s'écarte de deux conditions idéales: copie absolument fidèle du modèle ou, au contraire, copie entièrement remaniée pour se conformer à un système scripturaire particulier à une autre époque et/ou à une autre région. Tous les autres cas sont problématiques pour la fiabilité de la localisation, surtout s'il se superpose plusieurs couches de contamination au cours du temps:

... les textes [= manuscrits] examinés sont parfois des textes contaminés par suite de traductions interdialectales: une couche de formes récentes y coexiste avec une couche ancienne de formes primitives. [...] Ne trouvant aucun ensemble de chartes comparables au texte contaminé, le programme cherchera un compromis, c'est-à-dire un point d'enquête situé entre le point de provenance des formes primitives et le point de provenance des formes récentes. [...] une enquête supplémentaire sera nécessaire pour trouver le point géographique où les formes récentes ont été introduites dans les versions contaminées. (Dees 1999: 178)

C'est ainsi que les différents textes du ms. BnF 837, écrits de la même main, ont été localisés dans trois régions relativement différentes (cf. Dees 1987: xxix), ce qui laisserait croire, d'ailleurs, que l'intervention du copiste a été relativement discrète.

ms. BnF 837	Localisation automatique	
	région	coef.
<i>Marie l'Égypt.</i> Rutebeuf	Aisne-sud	(78)
<i>Théophile.</i> Rutebeuf	Yonne	(79)
<i>Vergi C</i>	Yonne	(80)
<i>Ailes de Raoul de Houdenc</i>	Yonne	(77)
<i>Songe d'enfer de R. de Houdenc</i>	Yonne	(77)
<i>Les dits de Hue Archevesque</i>	Eure	(83)
<i>Fabliaux</i>	Eure	(80)
<i>Trois aveugles de Compiègne</i>	Eure	(76)

Dees (1999) suggère un ensemble de correctifs permettant le cas échéant de distinguer deux couches dialectales dans la même copie. Ainsi pour les trois manuscrits suivants du *Charroi de Nîmes*, on pourrait aboutir à des localisations qui sont relativement distinctes de celles qui ont été attribuées par la localisation automatique:

manuscrits	Localisation automatique		Modèle	Copiste
	région	coef.		
frag F du <i>Charroi de Nîmes</i>	Langres	(93)	<i>id.</i>	<i>id.</i>
ms. A4 du <i>Charroi de Nîmes</i>	Nièvre, Allier	(79)	Langres	Tours, Chartres
ms. B1 du <i>Charroi de Nîmes</i>	Aisne	(85)	Langres	Paris, Soissons

On remarque ici, que même avec un coefficient de conformité avec la graphie des chartes de l'Aisne relativement élevé, puisqu'il est de 85, le manuscrit B1 du *Charroi de Nîmes* ne peut probablement pas servir comme témoignage dérivé. Selon les correctifs proposés, ses autres caractéristiques graphiques devraient se conformer tantôt à celles de la région de Langres, tantôt à celles de Paris ou Soissons. Il n'y a aucune raison de croire que la combinaison statistique des deux corresponde systématiquement à celles de l'Aisne.

On sait aussi qu'une source fréquente de la contamination des œuvres rimées provient de la difficulté qu'avaient les copistes à conserver les rimes de leurs modèles lorsqu'ils les rajeunissaient ou qu'ils les traduisaient dans un autre dialecte. Ils avaient, par conséquent, tendance à conserver les rimes originales sans en altérer les graphies. On obtiendrait un texte moins contaminé, dans ces cas, en ignorant les mots à la rime. Il est possible que cette technique ait été appliquée à certains des manuscrits de l'*Atlas des textes littéraires*, car on a retrouvé un des fichiers informatiques de *La Passion des Jongleurs* (éd. A.J.A. Perry, Paris 1981) dont tous les vers avaient été amputés des mots à la rime, peut-être en vue d'y appliquer l'algorithme de localisation.

3 Graphie et prononciation

3.1 Localisation: critères seulement graphiques

La localisation des manuscrits des textes littéraires se fonde exclusivement sur les usages graphiques. En effet, quelle que soit la régularité des correspondances phonographiques, seules sont pertinentes les graphies, jamais les prononciations qu'on pourrait en déduire. Ainsi l'utilisation du digraphe <lh> au XIII^e siècle dans les régions d'oïl est essentiellement une convention de l'espace culturel wallon et, en partie, hennuyer (cf. les cartes AC 8 et 199). Quelle que soit la valeur phonétique de ce digraphe, tout document qui le contient a plus de chance d'avoir été écrit en Wallonie ou dans le Hainaut qu'ailleurs.

Inversement, il est possible que la préposition 'à' se prononçait [e] ou [ɛ] en Bourgogne (tout comme la préposition contractées 'aux', cf. la carte AC 51), ce qui aurait alors constitué une propriété phonétique très caractéristique de cette région et des régions avoisinantes. Cette propriété n'aurait pourtant été d'aucun secours, puisque la tradition orthographique avait imposé le plus souvent, comme partout ailleurs, la graphie traditionnelle *a*.

3.2 Des systèmes scripturaires régionaux

Une fois le corpus rassemblé, avec un nombre suffisamment riche de témoignages fiables, primaires et dérivés, pour chacune des nombreuses régions du maillage géographique et idéalement aussi pour de nombreuses plages de temps, le dialectologue peut enfin s'atteler à la non moins formidable tâche de reconstruction des prononciations du passé.

À de nombreuses reprises, Dees annonce sa foi dans la possibilité de faire « parler » les graphies médiévales:

These considerations entail the conclusion that an attempt to reconstruct the regional diversity of spoken Old French should necessarily be based on an adequate description of the available written forms. Given the fact that written Old French differs systematically from one town to another, we will suppose here that the diversity observable on the level of the written language reflects, broadly speaking, a parallel diversity on the level of the phonetic forms. (Dees 1990:122)

Il faut cependant remarquer que si des différences graphiques d'une région à l'autre peuvent signaler des différences de prononciation, on ne peut jamais en être certain, sans autres éléments de preuves.

Le digraphe <lh> des formes wallonnes et hennuyères *ilh* de 'il' (cf. la carte AC 8) et *filh* de 'fils' représente certainement le même [ʎ] palatal que le <l> de *il* et de *fil* des régions avoisinantes de Champagne et de Lorraine.

L'usage du digraphe <oi> dans la graphie wallonne *dois* de 'deux' (cf. carte AC 103) présente le même cas de figure, même si dans ce cas le digraphe <oi> n'est pas spécifique aux chartes wallonnes. Rien ne serait plus erroné que de conclure que le cardinal 'deux' avait une prononciation différente en Wallonie où il s'écrit souvent *dois*, et dans les régions picardes et champenoises voisines où il s'écrit *deus*.

Le digraphe <oi> de la graphie wallonne *dois* de ‘deux’ note selon toute vraisemblance un son du type [ø], qui n’était pas nécessairement distinct de celui que note aussi dans cette région le digraphe <eu>, comme dans sa variante *deus* (cf. Goebel 2001: 841–842 pour d’autres usages de <oi>). On retrouve, beaucoup moins fréquemment, le même digraphe <oi> dans les graphies wallonnes de ‘seigneur’: *seignoir* (1264), *saignoir* (1248, 1266), *sanghoir* (1268), *sangoir* (1276). Les digraphes <oi> et <eu> ont pu se substituer l’un à l’autre après la convergence dans de nombreux parlers wallons des diphthongues issues de [e] et [o] roman accentué en syllabe ouverte vers un son commun, articulé [ø:] dans les parlers modernes, probablement selon les étapes suivantes: TRĒS > *treis* > *trois* > [trø:] et DŪŌS > *dous* > *deus* > [dø:] (cf. Francard 1980).

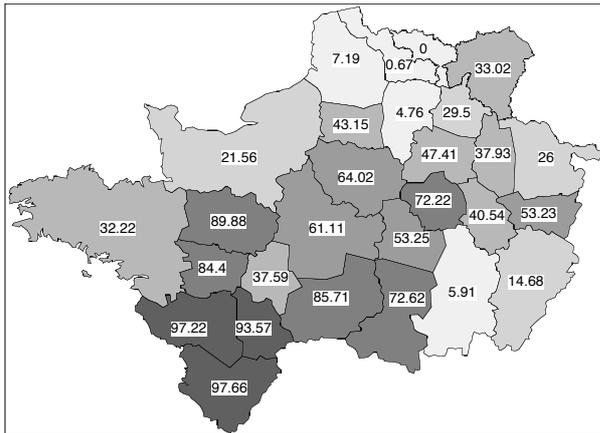
Il est significatif que malgré cette convergence phonétique, le remplacement de <eu> par <oi> (ou l’inverse) est relativement rare. La terminaison de ‘seigneur’ s’écrit en général *-or* ou *-our* en Wallonie (84% des cas, cf. AC 187, 188). Les graphies du type *saignoir* observées dès avant 1250, cependant, laissent croire que le changement vers un son de type [ø] s’était produit très tôt, et que les graphies *-or* et *-our* sont des usages anciens qui se sont maintenus malgré l’évolution phonétique. Les graphies de termes moins protocolaires que ‘seigneur’ sont peut-être plus susceptibles de s’adapter aux changements phonétiques. Pour ‘successeur’ (cf. AC 194), 53% des chartes utilisent la graphie *-eur*. L’usage relativement fréquent de <oi> dans la graphie de ‘deux’, sur le modèle de ‘trois’, s’explique certainement comme un effet de série, souvent observé avec les cardinaux.

Ainsi, ni les graphies en elles-mêmes, ni les différences de graphie entre des aires contigües, ne donnent la clé de la prononciation qu’elles signalent. Il est difficile d’imaginer, sans autre connaissance, que le digraphe <oi> du français moderne dans des mots comme *trois* note les sons [wa], ni qu’il indiquait la voyelle [ø] dans le même mot, écrit de la même façon, dans les chartes wallonnes du XIII^e siècle. Seul un patient travail de reconstruction permet de découvrir les sons qui se cachent sous les graphies.

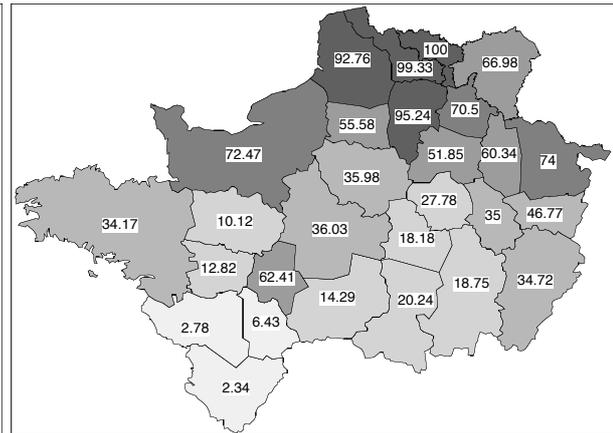
C’est ce que Dees reconnaissait, indirectement, dans son article de 1985, alors qu’il mentionne « l’existence, au XIII^e siècle, d’isoglosses qu’on retrouve telles quelles dans les dialectes du XX^e siècle » (1985: 112). En regrettant en même temps de n’avoir « rien pu faire jusqu’ici pour essayer de découvrir d’autres cas comparables à celui de la répartition géographique des variantes *as/aus* [de la préposition contractée ‘aux = à+les’, masculin et féminin] » qu’il avait abordé dans Dees (1971: 104–107), et sinon de renvoyer à une étude de Hilty (1973) pour un second cas de ce genre.

3.3 Les prépositions contractées *as/aus* (Dees 1971: 105–107)

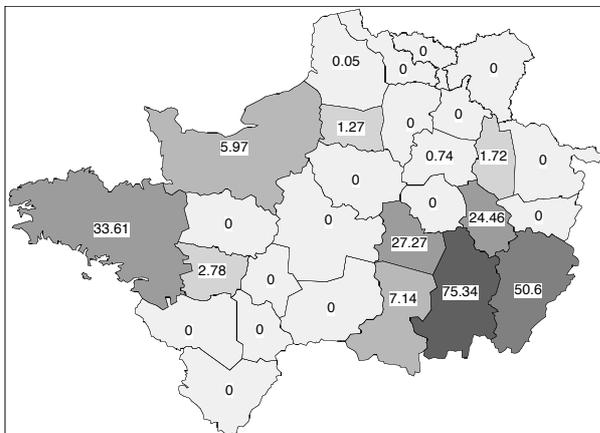
Examinons ce que nous apprennent les isoglosses qui délimitent les répartitions des variantes de la préposition contractée ‘aux = à+les’, en reprenant les données des cartes AC 50 et 51 de l’*Atlas des chartes*, qui n’avait pas encore été construit en 1971, et celles de la carte ALF 1122 ‘(jouer) aux quilles’ (la carte ALF 76 ‘aux autres’, utilisée par Dees 1971, n’ajoute rien à la précédente).



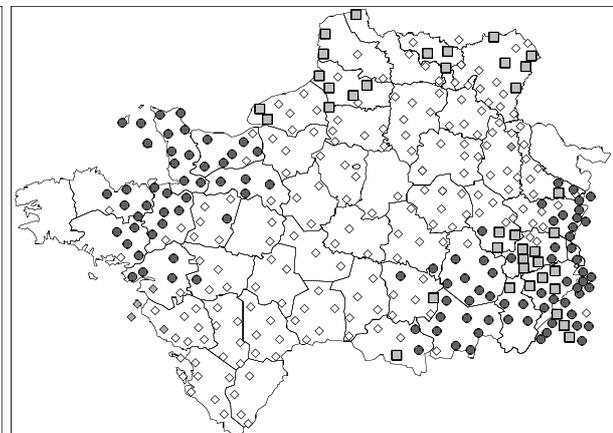
Carte 2a. 「aus」 ~ 「as, es」 de 'aux = à+les' – AC 50



Carte 2b. 「as」 ~ 「aus, es」 de 'aux = à+les' – AC 50/51



Carte 2c. 「es」 ~ 「aus, as」 de 'aux = à+les' – AC 50/51



Carte 2d. '(jouer) aux (quilles)' – ALF 112

Légende Carte 2d ◇ = [o] < [au] ◆ = [u] < [au] ■ = [a] < [a] ● = [e, ε] < [a]/[e]

La carte 2a reproduit la carte AC 50 de l'*Atlas des chartes*, tandis que les cartes 2b et 2c représentent, sous une forme différente, les données de AC 50 et 51. La carte 2a indique la proportion des graphies du type 「aus」 (= *aus, aux, auz, ax*) par rapport à l'ensemble des graphies du type 「aus, as, es」; et de même la carte 2b pour 「as」 (= *as, az*), et la carte 2c pour 「es」 (= *es, ez*). L'épicentre des graphies 「aus」 (carte 2a) se trouve dans la région Sud-Ouest ou, comme nous l'appellerons ici, le domaine poitevin (au sens de Avelle 2002: 479). Ces graphies s'étendent dans la direction nord-est où elles sont dominantes jusque dans la région parisienne. Le domaine des graphies 「as」 (carte 2b) a son épicentre dans le domaine picard restreint (sans l'Oise) et s'étend au sud-ouest en Normandie et au sud-est jusqu'en Lorraine. Enfin on observe (carte 2c) deux zones de part et d'autre d'un axe sud-ouest/nord-est pour les formes en 「es」: une importante, centrée sur la Bourgogne, où leur usage est fortement dominant et s'étend aux régions limitrophes, l'autre en Bretagne où l'usage s'en partage à égalité avec les deux autres variantes 「aus」 et 「as」, ne laissant voir que quelques traces en Normandie.

La carte 2d présente la répartition dialectale au début du XX^e siècle des prononciations de la même préposition contractée, telle qu'elle apparaît sur la carte ALF 112 'jouer *aux* quilles'. On voit une correspondance marquée entre les deux zones de la carte 2c où apparaissent les

graphies médiévales 「es」 et celles de la carte 2d avec les prononciations modernes [e], invitant à voir une filiation entre les usages phonétiques des deux périodes. Celle-ci exige cependant de faire un certain nombre d'hypothèses sur les changements qui se sont produits entre les XIII^e et XX^e siècles, si l'on admet qu'au XIII^e siècle les graphies 「aus」, 「as」 et 「es」 notaient respectivement des prononciations du type [a_u ~ a_uz], [a ~ az] et [e ~ ez].

Le résultat [e] dans la première zone — en Bretagne et dans la partie occidentale de la Normandie — serait l'aboutissement d'un changement phonétique plus ou moins variable [a_u] > [a] > [e], dont les trois étapes seraient également représentées dans les graphies des chartes médiévales bretonnes de cette période, mais dont la dernière étape aurait à peine été commencée en Normandie. Dans la seconde zone, la correspondance entre les 「es」 médiévaux et les formes modernes [e] est étroite en Bourgogne, tandis que plus à l'Est, en Franche-Comté et surtout dans les régions frontalières de Lorraine, le passage de [a] > [e] serait plus tardif, laissant survivre de nombreuses formes [a], contrairement à ce qui s'est produit à l'Ouest.

Il est probable aussi que la nature des changements historiques en œuvre n'est pas la même dans ces deux zones. Le passage de [a] > [ɛ] > [e] s'observe dans un grand nombre d'autres contextes à l'Est et affecte en particulier la préposition simple 'à'. Ce qui n'est pas le cas en Bretagne, ni en Normandie, où Dottin et Langouët (1901: cvii) suggèrent — à défaut d'un changement phonétique plausible — qu'un croisement a pu se produire avec la préposition contractée 'es = en+les', ce qui n'est pas très satisfaisant non plus.

Les reflets [o] de [a_u] se sont étendus bien au-delà du domaine médiéval de 「aus」. On peut arguer que dans l'Orléanais et dans la région parisienne, [o] est bien le continuateur phonétique d'un ancien [a_u], représenté par les graphies 「aus」 majoritaires (61 % et 64 %), qui s'est généralisé au détriment du [a] écrit 「as」 avec lequel il variait. La forme [o] dans les dialectes du domaine picard et de la partie orientale de la Normandie aurait une toute autre origine. Si la prononciation médiévale était bien [a] comme l'indique la graphie dominante, la nouvelle prononciation [o] s'expliquerait par la pénétration, par emprunt, du [o] de la norme parisienne, puisque ces régions ne connaissent pas le passage de [a] *non accentué* > [o], et en particulier que la préposition simple 'à' y est demeurée [a]. Il est de plus significatif que la variante moderne [a] n'est pas rare dans les régions normandes, wallonnes et picardes (à l'exception de l'Aisne), où elle perpétue la prononciation médiévale [a] des graphies 「as」.

Si l'on admet ces hypothèses — raisonnables — de changement, alors on peut certainement conclure avec Dees (1985:112) à « l'existence, au XIII^e siècle, d'isoglosses qu'on retrouve [presque] telles quelles dans les dialectes du XX^e siècle » (on me permettra d'ajouter *presque* avant *telles quelles*.) Des recherches ultérieures devraient cependant examiner le passage de [a] à [e] ou à [o], sur lequel on est très mal informé. Ainsi, des changements morphologiques et syntaxiques ont profondément modifié le statut des continueurs de la préposition contractée médiévale *as* 'aux' en Picardie, dont il faudrait tenir compte pour expliquer les distributions modernes. Notons que le déterminant 'le, la, les' dans de nombreux parlers picards est limité à des constructions plus ou moins spécifiques et a été remplacé ailleurs par les continueurs du démonstratif *che/chel, chelle, ches*. La préposition contractée *as* a non seulement été remplacée par des constructions syntaxiques régulières du type 「à ches」, mais parfois aussi 「à les」. Elle a aussi pu être refaite sur la forme du masculin singulier, qui s'écrivait *au* au moyen âge (cf. AC 48 et 49), ou du féminin [al] issu des formes médiévales écrites *a le*.)

Nous avons vu comment, dans ce cas précis, on a pu concilier l'existence d'isoglosses modernes et la variabilité des graphies dans les chartes de nombreuses régions, en supposant que la variabilité graphique reflétait alors celle de la prononciation. Plus généralement, voici des raisons qui peuvent être avancées pour la variabilité des graphies dans une région donnée:

1. il s'agit d'« accident[s] dont il est permis de ne pas tenir compte » (Dees 1971:106),
2. il s'agit d'« échanges inévitables » « près de la limite qui sépare deux aires » (Dees, *id.*),
3. la prononciation et la graphie sont variables dans cette région,
4. la graphie, mais non la prononciation, est variable dans cette région, certaines variantes graphiques étant des reliques d'un usage scripturaire plus ancien,
5. la graphie, mais non la prononciation, est variable dans cette région, certaines variantes graphiques étant empruntées à un usage scripturaire extérieur à la région.

4 Dees, la « scripta » et la langue parlée au moyen âge

Dees s'est toujours opposé à une forme particulière de la dernière hypothèse, celle qui verrait dans une partie au moins de la variabilité graphique des chartes, l'influence de la pratique scripturaire de la région parisienne. Ceci l'a amené à s'attaquer fortement (*in* Dees 1985) au travail de Remacle (1948), qui aurait été à l'origine de la *scriptologie*, un courant de recherche qui voit dans les graphies régionales du XIII^e siècle l'expression d'un « français unitaire » où les « formes franciennes [...] se sont imposées » et où apparaissent dans une proportion plus ou moins importante des traits graphiques « nés sous l'action d'un substrat dialectal générateur » (Goebel 2001: 844, 849).

Le terme *scripta*, défini à l'origine comme « la langue vulgaire écrite au moyen âge » (Remacle 1948: 15) et qui pouvait être spécifique à chaque région donnée, apparaît dans le travail de Dees avec le sens particulier de « scriptologie », semble-t-il. Dans la critique qu'il en fait, Dees rejette l'existence d'une « koinè écrite pour la période de l'ancien français » et d'un « français écrit commun [qui] serait en train de se répandre sur le domaine d'oïl dès avant 1300 » (Dees 1987: vii). Ce qui l'amène à conclure, « en ce qui concerne la période antérieure à 1300, [à] l'entière fausseté de la notion de scripta » (1985:111–112).

Si la critique que fait Dees de l'influence imputée à Paris sur les usages scripturaire des régions d'oïl pendant et avant le XIII^e siècle semble bien souvent justifiée, son argumentation ne permet cependant d'exclure aucune des cinq hypothèses précédentes.

5 Des normes régionales

En particulier, on ne peut exclure que certaines graphies aient été empruntées à des aires culturelles extérieures, comme il apparaîtra des deux cas suivants.

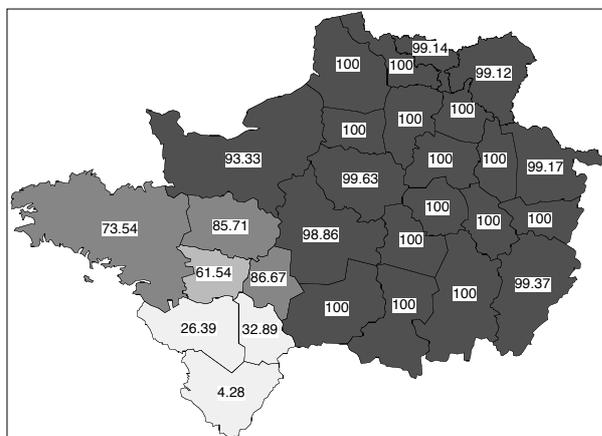
5.1 La graphie <ou> pour [ɔ] dans le domaine liégeois

Les cartes 1a et 1b « 「ou」 ~ 「o」 de l'indéfini 'tout, tous, toute(s)' » montrent la progression du digraphe <ou> au détriment de <o> dans le dernier quart du XIII^e siècle en Wallonie. La même progression s'observe dans les chartes propres à la Principauté de Liège, où les graphies <ou>

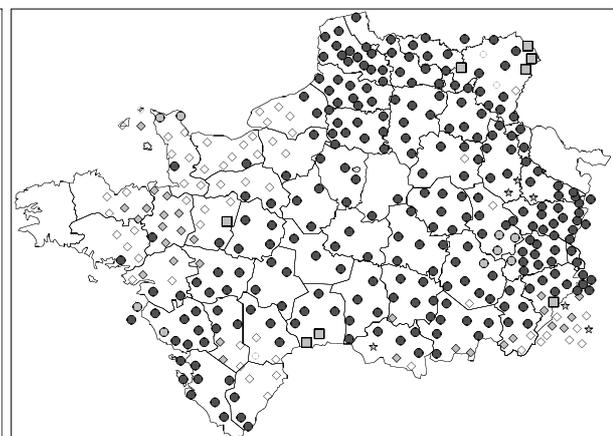
passent de 8% (pour 12 chartes avant 1275) à 45% (pour 5 chartes après cette date). Or c'est une caractéristique du wallon liégeois que la voyelle romane [o] accentuée entravée ne s'est généralement pas fermée, comme dans la plupart des parlers d'oïl, mais au contraire qu'elle s'est ouverte devant les obstruantes, et en particulier dans les formes de ce mot: [tɔ] 'tout, tous' et [tɔt] 'toute, toutes' (cf. ALW1: 123, 201 et Haust 1948, qui n'enregistre aucune variation de prononciation dans le domaine liégeois).

Il n'y a donc aucune raison pour que la graphie passe de <o> à <ou> dans ce mot, quelles que soient celles qu'on peut invoquer pour son évolution ailleurs (cf. Dees et van Reenen 1980). On peut certainement y voir l'adoption d'une norme scripturale provoquée par les échanges écrits qu'accompagnait une intense activité commerciale et légale des Liégeois avec leurs voisins picards: les scribes liégeois auront lu [ɔ] ce qu'ils voyaient écrits *ou* et auraient transposé cet usage dans leur pratique. C'est peut-être ce type d'évolution des usages graphiques que Dees classait sous la rubrique des « échanges inévitables », même s'il a pu se produire dans le domaine liégeois au-delà de la limite qui le sépare de l'aire hainuyère; cette démonstration demanderait cependant des données plus riches. (On ne peut exclure non plus l'influence combinée de graphies anciennes <ou> pour [ɔ:] long, comme celle du mot 'coup', dont l'évolution dans le domaine liégeois a pu être la suivante: cōLĀPŪM > [kɔlp] > [kɔpp] > [kɔ:p] > [ko:(p)], si on admet que la perte du [l] préconsonantique après [ɔ] s'est faite après une étape de vocalisation notée <u> dans les graphies médiévales.)

5.2 La graphie <oi> pour [ɛ] en Normandie et en Bretagne



Carte 3a. 「oi」 ~ 「ei」 de 'savoir' – AC 255



Carte 3b. 'savoir' – ALF 1200

Légende Carte 3b ◇ = [e, ε] < [eṽ] ◆ = [a] < [eṽ] ■ = [ø] < [oṽ], sauf dans le Maine
 ● = [ɥe, ɥa] < [oṽ] ☆ autre suffixe ● = [o] < [oṽ], sauf dans le Cotentin

Les voyelles ◇ et ◆ sont issues de [eṽ] primitif sans passer par une étape [oṽ].

Le résultat [ø] (ou [ə]) dans le Maine provient d'une centralisation plus tardive de [e] accentué < [eṽ].

Le résultat [o] au nord du Cotentin est le traitement normal de *-er* < *-ĀRĒ*, avec lequel *-eir* s'est confondu.

On examine ici les variantes graphiques médiévales 「oi」 ~ 「ei」 notant les voyelles issues de [eṽ] de l'ancien français primitif et les correspondants modernes de ces voyelles. S'il y avait un air de parenté entre les distributions de la préposition contractée 'aux = à+les' des cartes de

l'*Atlas des chartes* et celles de l'ALF, aucune correspondance régulière n'apparaît dans ce cas, comme on voit sur les cartes 3a et 3b pour la terminaison de l'infinitif du verbe 'savoir'.

Les divergences les plus marquées s'observent dans la partie ouest du domaine d'oïl, où apparaît une *inversion* radicale. Le domaine poitevin où dominaient les résultats ^ɾei (zones claires de la carte 3a) a maintenant des formes qui proviennent d'un ancien [oi], tandis qu'en Bretagne romane, dans le Maine, en Anjou, en Touraine et surtout en Normandie (zones plus foncées de la carte 3a), c'est la situation inverse.

Les formes modernes du domaine poitevin ne sont pas vraiment problématiques. Elles s'expliquent en partie par la pénétration des formes standard dans les dialectes, et en partie par une évolution propre à certaines régions, où [ei] > [oi] plus tardivement (et qui affecte aussi des diphtongues [ei] de formation plus récente, comme dans [groɛl] 'grêle'). Par contre, les formes modernes des autres régions de l'Ouest, en particulier celles de la Normandie à laquelle se limiteront maintenant nos observations, soulèvent de sérieuses questions sur la relation entre les graphies médiévales du type ^ɾoi et la prononciation.

Le [ei] primitif ne peut pas être passé par une étape [oi] en Normandie (sauf dans la partie orientale près du domaine picard). L'hypothèse d'un changement du type: afr. prim. [ei] > [oi] > [ɥe] > [e], n'est pas admissible dans ces régions, car le même changement n'affecte pas les [oi] primitifs, par exemple: NŪCĒM > afr. [noɪz] > [nɥe] 'noix', et non pas *[ne] (cf. Chauveau 1989, Schogt 1960). En l'absence d'une étape [oi] dans cette région, comment expliquer que les chartes utilisent régulièrement le digraphe <oi> pour représenter les reflets du [ei] primitif? Cette graphie problématique n'est pas spécifique à l'infinitif de 'savoir', mais vaut pour l'ensemble des formes communes de l'*Atlas des chartes* et de l'ALF (aussi bien les formes verbales, AC 224 – ALF 517 (*qu'il soit*, AC 227 – ALF 95 (*il avait*, AC 228 – ALF 510/511 (*il était*, AC 247 – ALF 82/588 *avoir*, que les autres: AC 104 – ALF 1333 *trois*, AC 146 – ALF 428 *droit*, AC 177 – ALF 888 *mois*). Il n'y a que le mot *hoir/heir* 'héritier' (AC 158), que nous examinerons plus bas, dont la distribution des variantes médiévales ^ɾoi ~ ^ɾei est plus conforme à l'évolution dialectale attendue du [ei] primitif.

La réponse semble inéluctable: les scribes autochtones écrivaient *oi* en Normandie, mais lisaient [e(i)] ou [ɛ(i)]. Le digraphe <oi> semble bien avoir été emprunté à des pratiques scripturaires extérieures. Lesquelles? Il est peu vraisemblable que ce soient celles de la région parisienne où la même question se pose, comme nous allons voir.

5.3 La graphie <oi> pour [ɛ] dans la région parisienne

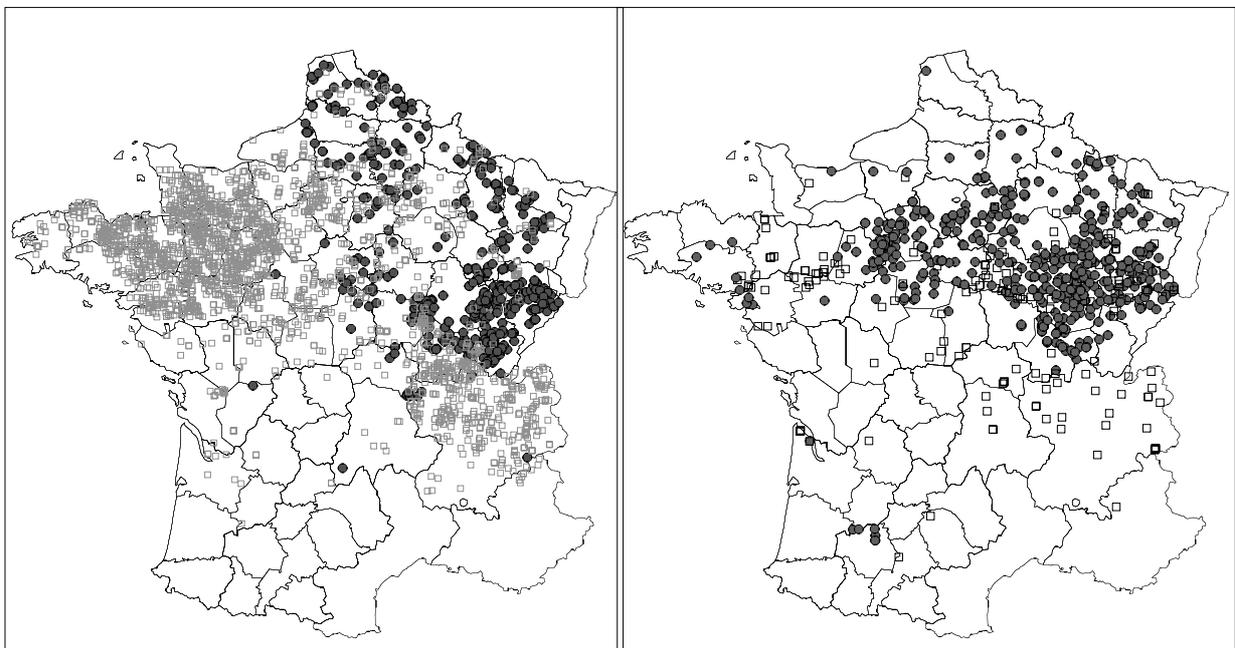
Paris n'a pas été à l'avant-garde de l'emploi du vernaculaire dans les écrits légaux, le pouvoir royal ayant longtemps favorisé l'usage du latin dans sa chancellerie (cf. Lusignan 2004). Les clercs parisiens ont très bien pu s'inspirer des usages scripturaires d'autres régions lorsqu'ils ont commencé à rédiger des actes en français et interpréter <oi> comme une graphie pour [e(i)] ou [ɛ(i)]. L'étude des rimes dans des manuscrits parisiens laisse clairement voir cette possibilité.

Les usages graphiques du ms. BnF fr. 784 du *Roman de Thèbes*, du ms. L de la *Chastelaine de Vergi* et du ms. de *Ste-Catherine d'Alexandrie* sont conformes à ceux des chartes de la région parisienne (cf. Dees 1987:527 pour la localisation) et utilisent des graphies fort diverses pour noter les sons [e(i)] ou [ɛ(i)] assurés par les rimes: *voire* 'vérité': *affaire* (*Rom.*

Thèbes, ms. BnF fr. 784), *tricherresse* ‘tricheuse’: *duchoisse* ‘duchesse’, *dueil* ‘je souffre’: *voil* ‘je veux’ (*Chastelaine de Vergi*, ms. L), *essaient*: *soient*, *poine*: *regne*, *aies*: *soloies*, *feite*: *destroite*, *deull*: *vuouill* (aussi écrit *voill* ailleurs) (*Ste-Catherine d’Alexandrie*), où non seulement les digraphes <oi> et <ai> se lisaient ainsi dans ces rimes, mais <ue> et <eu>. (Il ne reste de trace de cette prononciation de <ue> et <eu> dans la langue standard que dans le mot *bienveillant*, pour *bienvueillant*; au XVI^e siècle, par contre, la lecture [e] ou [ɛ] de <ue> et <eu> était encore très fréquente).

On sait d’ailleurs que <oi> continuera à être une graphie possible à Paris pour le son [ɛ], comme dans les mots *françois*, *froid*, (*qu’il*) *soit*, (*il*) *savoit* — comme le rapportent les grammairiens du XVII^e siècle qui ne condamnent en général plus cet usage, contrairement à ceux du XVI^e, pour qui [ɛ] était, selon les uns ou les autres, une prononciation « vulgaire » ou « courtisane ».

À défaut de pouvoir s’appuyer avec certitude sur les graphies des chartes du XIII^e siècle, on peut faire appel — avec prudence — aux graphies modernes de certains toponymes pour estimer quelle a été l’extension maximale du passage de [eɪ] primitif à [oɪ] avant le XIII^e siècle. Comme le montre Haudricourt (1947, 1972), les graphies modernes des toponymes construits sur des noms dérivés des noms d’arbre en -ĒTŪM, -ĒTĀ, comme *Aulnais/Aulnois* ou *Aulnaye/Aulnoye* peuvent donner une idée des distributions géographiques des formes anciennes. Il faut bien sûr tenir compte du figement possible des graphies à l’étape ‘ei’ (écrites *ei/ey*, *ai/ay* ou même *é*) dans des régions où ‘oi’ a fini par prévaloir et qu’inversement les usages qui faisaient écrire *oi/oy* le son [e(i)] ou [ɛ(i)] dans les chartes ont aussi pu se faire sentir: la graphie *Rozoy*, longtemps utilisée malgré la prononciation locale [rozɛ] a fini par être remplacée en 1934 dans l’appellation de *Rozay-en-Brie* (au centre de la Seine-et-Marne).



Carte 4a. Toponymes Aunoy / Vernoy

Carte 4b. Toponymes Charmoy / Ormoy

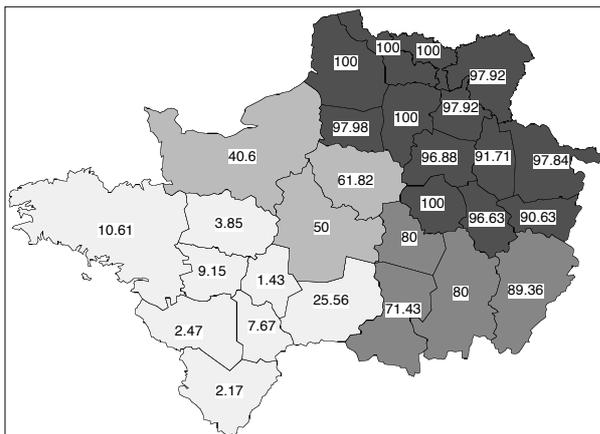
Les cartes 4a et 4b ont été construites à partir des toponymes portés sur les cartes IGN au 1: 25 000 qui permettent d’avoir un très grand nombre d’attestations (plus de 2500 pour la première, environ 650 pour la seconde). Les points noirs notent les résultats du type ‘-oi’ (*-oi*, *-oie*, *-oy*, *-oye*, *-ois*) et les carrés blancs (qui forment un véritable nuage gris sur la carte 4a)

ceux du type «*-ei*» (*-ai, -aie, -ay, -aye, -ais, -é, -ée, -és*). Les toponymes en *-et* qui peuvent se confondre avec des diminutifs *-et* < *-ĭttŭm* n'ont pas été inclus.

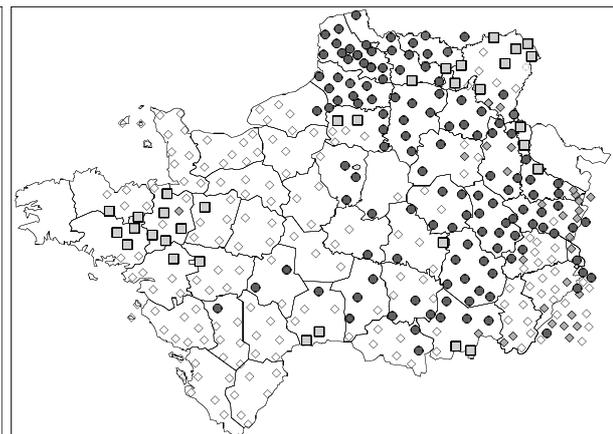
La carte 4a consigne les résultats de ^oALNĒTŪM, ^oALNĒTĀ ‘lieu planté d’aulnes’ (*Aunoy, Aulnay, Launay, Lannoy, Les aunaies*, etc.) et de ^oVERNĒTŪM, ^oVERNĒTĀ, ayant le même sens (*La Vernaie, Verney, Les Verneys, Vernoy*, etc.); elle permet de voir l'évolution du [eĭ] primitif après une dentale. La carte 4b consigne ceux de ^oCARPĪNĒTŪM, ^oCARPĪNĒTĀ ‘lieu planté de charmes’ (*Charmoi, Charmaie, Carmoy, Carmay*) et de ^oŪLMĒTŪM, ^oŪLMĒTĀ ‘lieu planté d’ormes’ (*Ormoie, Ourmoi, L'houmaie, L'hommée*, etc.); elle permet de voir l'évolution de la même diphtongue après une labiale. La densité bien moindre des toponymes de la seconde carte vient probablement de facteurs sociaux et environnementaux, mais aussi linguistiques: en particulier, la suite [-pn-] < -PĪN- dans ^oCARPĪNĒTŪM donne le plus souvent [-n-] dans le nord du domaine d'oïl, d'où des toponymes du type *Carnoy* et *Charnaie*, qui n'ont pas été retenus.

Ces deux cartes confirment le résultat bien connu que l'évolution du [eĭ] primitif est sensible à la nature de la consonne précédente. Les parlers ruraux du centre de la France, où les différences sont les plus marquées, n'ont pas participé aussi intensément que la langue standard à l'évolution [eĭ] > [oi] après une consonne non labiale. On remarque en particulier que les toponymes en «*oi*» — comme *Aulnoy* ou *Vernoy* — sont remarquablement absents de l'Île-de-France et de l'Orléanais (sauf, respectivement, dans la Brie et dans la partie orientale du Loiret).

La plupart des cartes de l'ALF ne laissent cependant pas transparaître cette distinction selon la consonne précédente — sauf, en partie, dans le domaine poitevin. Dans les autres régions, les distributions des continueurs du [eĭ] primitif (cf. ALF 916 ‘noir’ et 1238 ‘soir’) sont relativement semblables à celle qui apparaît sur la carte 3b (ALF 1200 ‘savoir’) et non à celle de la carte 4a (toponymes *Aulnoy* ou *Vernoy*) — ce qui peut s'expliquer par l'influence de la langue standard.



Carte 5a. «*hoir*» ~ «*heir*» – AC 158



Carte 5b. «*froid*» – ALF 612

Légende Carte 5b ◇ = [e, ε] < [eĭ] ◆ = [a] < [eĭ]
 ● = [ʏe, ʏa, o] < [oi] ■ = [ø] < [oi], sauf Bretagne, Maine et Anjou

Les voyelles ◇ et ◆ sont certainement issues de [eĭ] primitif sans passer par une étape [oi].

Le résultat [ø] ([ə]) en Bretagne, dans le Maine et en Anjou provient d'une centralisation plus tardive.

Il existe néanmoins un petit nombre de mots relevés dans l'ALF, pour lesquels la distribution des reflètes du [eĭ] primitif en initiale de mot ou après une consonne non labiale est relativement conforme à son évolution dans les toponymes. Ce sont les mots *froid* (ALF 612), *droit* (ALF 428), *droite* (ALF 427), *endroit* (ALF 460), *raie* (ALF 1234), il *fallait* (ALF 536) et plus généralement toutes les formes du singulier de l'imparfait et du conditionnel des verbes, ainsi que les témoignages isolés de *toit* (ALF 451, 1310). Leur distribution est dans l'ensemble voisine de celle qui a été portée ici sur la carte 5b pour *froid* et pourrait refléter un développement autochtone (on notera cependant une nette dominance de formes où l'ancienne diphtongue était précédée de [r], ce qui pourrait aussi expliquer en partie ces résultats).

La distribution de 「oi」 ~ 「ei」 pour *hoir/heir* 'héritier' de l'*Atlas des chartes* (AC 158), présentée ici sur la carte 5a, rappelle aussi celle des toponymes *Aulnoy* et *Vernoy* (carte 4a) et de *froid* (carte 5b) — plutôt que celle de 'savoir' (carte 3a), bien que la proportion des graphies 「oi」 reste relativement élevée en Normandie, dans l'Orléanais et dans la région parisienne.

Ce faisceau d'observations laisse supposer que le changement [eĭ] > [oi] a pu être relativement variable avant de se stabiliser dans la région parisienne et dans l'Orléanais. Cette variabilité a pu favoriser une certaine incertitude dans l'usage des graphies <oi> ou <ei>. Elle ne saurait expliquer, seule, l'emploi massif du digraphe <oi> dans les chartes de ces deux régions pour une grande majorité des mots où [eĭ] primitif n'était pas précédé d'une consonne labiale. Pour 'trois' (AC 104) et 'droit' (AC 146), le pourcentage des variantes 「oi」 dépasse toujours 90%. Il faut probablement y voir un facteur culturel semblable à celui qui est responsable de l'usage de ce même digraphe <oi> en Normandie, où il s'est imposé malgré une prononciation dominante de type [e(i)] ou [ɛ(i)].

C'est au nord et à l'est du domaine d'oïl que le digraphe <oi> avait un support phonétique indéniable à l'origine. La riche et commerçante Picardie est la région la plus susceptible d'avoir marqué de son empreinte graphique les échanges écrits sur ce vaste domaine.

L'homogénéité — toute relative — de la graphie des lettrés n'impliquait cependant pas une normalisation des prononciations, comme cela se produira au XVI^e siècle. Les graphies propagées par l'imprimerie tendront à avoir cet effet et serviront de cadre pour l'établissement d'une prononciation uniforme à l'usage des classes dominantes, avec un certain succès, il faut le reconnaître. Elles se montreront néanmoins incapables de canaliser tous les usages parisiens ordinaires, dont certaines formes survivent dans la norme moderne: les terminaisons de l'imparfait *-ais*, *-ait*, *-aient*, entre autres, mais aussi *français*, *anglais*, *marais*, *monnaie*, *emplette*, etc.

6 En conclusion

Une dialectologie précise de l'ancien français exige un examen rigoureux des documents anciens permettant d'identifier ceux qui sont les plus représentatifs des usages des différentes régions à des moments précis dans l'histoire. Ceci implique dans un premier temps l'identification des caractéristiques de documents originaux qu'on a pu dater et localiser de

façon précise, appelés « témoignages primaires »; puis, à partir de ces caractéristiques, le classement et l'interprétation des autres documents. C'est à cette tâche formidable que s'est attaquée l'équipe d'Amsterdam sous la direction de Anthonij Dees avec une efficacité remarquable.

Cette tâche s'appuie entièrement sur l'examen des graphies (cf. Dees 1991:700). Les résultats publiés dans l'*Atlas des chartes* montrent sans conteste que les usages scripturaires des différentes régions d'oïl sont suffisamment individualisés pour qu'on espère pouvoir trouver un ensemble de traits discriminants qui permettrait d'identifier rigoureusement les documents originaux de chaque ville, comme le postulait Dees (1990:122) — à condition d'avoir un nombre suffisant de témoignages primaires pour chacune d'entre elles. Les procédures de localisation des manuscrits produisent une mesure objective de la conformité des usages graphiques d'un document avec ceux d'une région donnée. Et même lorsque les conditions idéales sur le nombre des témoignages primaires ne sont pas totalement remplies, cette mesure donne un indice suffisant pour relativiser la teneur des hypothèses que l'on serait tenté de faire sur l'état et l'évolution des dialectes.

En ce qui concerne l'interprétation phonétique des graphies du XIII^e siècle, cependant, la recherche n'a été qu'ébauchée. On ne peut ni dériver directement les prononciations des graphies, ni supposer que des différences entre les usages graphiques de deux régions signalent des différences de prononciation. Par exemple, l'isoglosse entre le Département du Pas-de-Calais et le Département du Nord qui divise la Picardie selon l'usage qu'on y fait des graphies <ch> et <c> de l'initiale du démonstratif 'ce' (AC 27) et du cardinal 'cent' (AC 117) ne correspond certainement pas à des différences de prononciation.

Les graphies des documents du XIII^e siècle sont un élément, certes important, mais seulement un élément parmi tous ceux qui sont nécessaires à la reconstruction des prononciations. On doit également examiner les données en aval, en particulier les reflets modernes de la diversité dialectale médiévale, comme Dees avait commencé à faire en 1971. Et les données en amont: en particulier, déterminer comment se sont développés, institués, et propagés les systèmes scripturaires. Ces systèmes ont toujours été sous la dépendance du système graphique du latin, tel qu'il était compris et enseigné dans chaque région. Tant que les transcriptions graphiques des vernaculaires n'étaient pas « institutionnalisées », les valeurs phonétiques des lettres utilisées ne pouvaient que correspondre aux valeurs phonétiques régionales qu'on leur donnait dans l'enseignement du latin. Ainsi <ei> devait à l'origine représenter une diphtongue du type [eī] ou [ēī] et <oi> une diphtongue du type [uī], [ōī] ou [ōī]. À partir du moment où la pratique écrite du vernaculaire devient courante, certaines combinaisons de lettres peuvent se figer dans les graphies des mots, indépendamment de leurs prononciations. Si dans les années 1250, des clercs wallons utilisaient occasionnellement, sur le modèle de *trois*, le digraphe <oi> pour rendre la voyelle tonique de 'seigneur', c'est que la graphie <oi> dans ces deux mots devait déjà noter un son du type [ø]. Il n'est pas impossible, au moins pour certains scribes, que le mot 'seigneur' ait eu une prononciation unique et qu'ils l'écrivaient occasionnellement *seignoir*, mais plus normalement *sangnor*, *sangnour* ou *sangnur*, selon les usages graphiques conventionnels transmis dans les scriptoria ou ailleurs. Mais quoi qu'il en soit de la prononciation variable ou non de ces mots, l'usage occasionnel de <oi> pour [ø] en Wallonie indique que la pratique de l'écriture du vernaculaire y est courante dans certains milieux, et qu'il existe alors un système scripturaire en marge du latin qui fixe en partie les graphies des mots du vernaculaire, indépendamment de leurs prononciations.

Références

AC = Dees (1980).

ALF = Gilliéron et Edmont (1902–1910).

ALWI = Remacle (1953).

Avalle, D'Arco Silvio (1962): *Cultura e lingua francese delle origini nella « Passion » di Clermont-Ferrand*. Milan: Riccardo Ricciardi. [Repris en 2002 dans *Fenomenologia ecdotica et lingua letteraria del medioevo romanzo*, éd. par Lino Leonardi, p. 441–549. Firenze: Edizioni del Galluzzo e Fondazione Ezio Franceschini].

Bonnier, Charles (1889–1890): Étude critique des Chartes de Douai de 1203 à 1275. *Zeitschrift für romanische Philologie* 13–14.431–462, 14.66–88, 298–343.

Chauveau, Jean-Paul (1989): *Évolutions phonétiques en gallo*. Paris: CNRS.

Dees, Anthonij (1971): Étude sur l'évolution des démonstratifs en ancien et en moyen français. Groningen: Wolters-Noordhoff.

Dees, Anthonij, avec le concours de Pieter van Reenen et Johan de Vries (1980): *Atlas des formes et des constructions des chartes françaises du XIII^e siècle*. Tübingen: Niemeyer.

Dees, Anthonij (1985): Dialectes et scriptae à l'époque de l'ancien français. *Revue de linguistique romane* 49.87–117.

Dees, Anthonij, avec le concours de Marcel Dekker, Onno Huber et Karin van Reenen-Stein (1987): *Atlas des formes linguistiques des textes littéraires de l'ancien français*. Tübingen: Niemeyer.

Dees, Anthonij (1990): Towards a dialectology of spoken Old French: the analysis of rhymes. *Historical linguistics and philology*, éd. par Jacek Fisiak, 119–134. Mouton: De Gruyter.

Dees, Anthonij (1991): Ancien français écrit et ancien français parlé. *Actes du XVIII^e Congrès International de Linguistique et Philologie Romanes*, vol. 3.700–705. Tübingen: Niemeyer.

Dees, Anthonij (1999): La tradition manuscrite du *Charroi de Nîmes*. *La Recherche: Bilan et perspective. Le moyen français* 44–45.129–189.

Dees, Anthonij (2001): La dialectologie historique. *Lexikon der romanistischen Linguistik*, vol. 1.2, éd. par Günter Holtus, Michael Metzeltin et Christian Schmitt, pp. 852–856. Tübingen: Niemeyer.

Dees, Anthonij et Pieter van Reenen (1980): L'interprétation des graphies -o- et -ou- à la lumière des formes trouvées dans les chartes françaises du XIII^e siècle. *Linguistic studies offered to Berthe Siertsema*, éd. par D. J. Alkemade, A. Feitsma, W. J. Meys, P. van Reenen et J. J. Spa, 269–275. Amsterdam: Rodopi.

Dottin, Georges et J. Langouët (1901): *Glossaire du parler de Pléchéâtel (canton de Bain, Ille-et-Vilaine)*. Rennes: Plihon et Hommay.

Fallot, Gustave (1839): *Recherches sur les formes grammaticales de la langue française et de ses dialectes*, publiées par Paul Ackermann et précédées d'une notice sur l'auteur par M. B. Guérard. Paris: Imprimerie Royale.

Francard, Michel (1980): *Le parler de Tenneville. Introduction à l'étude linguistique des parlers wallon-lorrains*. Louvain-la-Neuve: Cabay.

Gilliéron, Jules et Edmond Edmont (1902–1910): *Atlas linguistique de la France*. Paris: Champion.

Goebel, Hans et Jakob Wüest. 2001. *Scriptologie*. *Lexikon der romanistischen Linguistik*, vol. 1.2, éd. par Günter Holtus, Michael Metzeltin et Christian Schmitt, pp. 835–851. Tübingen: Niemeyer.

Haudricourt, André-Georges (1947): Français *ei* > *oi*. *Lingua* 1.209–218. [Repris en 1972 dans *Problèmes de phonologie diachronique*, pp. 57–72. Paris: Sela.]

- Haupt, Jean (1948): Dictionnaire français-liégeois. Liège: Vaillant-Carmanne.
- Hilty, Gerold (1973): Les origines de la langue littéraire française. *Vox Romanica* 32.254–271.
- Huber, Onno et Karin van Reenen-Stein (1988): Corrélations et regroupements dans l'*Atlas des formes et des constructions des chartes françaises du XIII^e siècle*. Distributions spatiales et temporelles, constellations des manuscrits. Études de variation linguistique offertes à Anthonij Dees à l'occasion de son 60^e anniversaire, éd. par Karin van Reenen-Stein et Pieter van Reenen, 93–101. Amsterdam: John Benjamins.
- Lusignan, Serge. 2004. La langue des rois au Moyen Âge – Le français en France et en Angleterre. Paris: Presses Universitaires de France.
- Philipon, Epaminondas (1910): Les parlers du Duché de Bourgogne aux XIII^e et XIV^e siècles. *Romania* 39.476–531.
- Philipon, Epaminondas (1912): Les parlers du Duché de Bourgogne aux XIII^e et XIV^e siècles, 2. La Bourgogne occidentale. *Romania* 41.541–600.
- Philipon, Epaminondas (1914): Les parlers de la Comté de Bourgogne aux XIII^e et XIV^e siècles. *Romania* 43.495–559.
- Remacle, Louis (1948): Le problème de l'ancien wallon. Liège: Faculté de Philosophie et Lettres.
- Remacle, Louis (1953): Atlas Linguistique de la Wallonie, tome 1: Aspects phonétiques. Liège: Vaillant-Carmanne.
- Schogt, Henry Gilius (1960): Les causes de la double issue de *e* fermé tonique libre en français. Amsterdam: Van Oorschot.
- Wailly, Natalis de (1870): Recueil de chartes en langue vulgaire provenant des archives de la collégiale de Saint-Pierre d'Aire. *Bibliothèque de l'École des Chartes* 31.261–302.
- Wailly, Natalis de (1871): Observations grammaticales sur des chartes françaises d'Aire en Artois. *Bibliothèque de l'École des Chartes* 32.291–320.
- Wailly, Natalis de (1876): Observations sur la langue de Reims au XIII^e siècle. *Mémoires de l'Académie des Inscriptions et Belles Lettres* 28:2.287–316.
- Wilmotte, Maurice (1888): Études de philologie wallonne. *Romania* 17.542–590. Repris en 1932 dans *Études de philologie wallonne*, p. 41–150. Paris: Droz.

Remerciements

Ce n'est pas sans émotion que je tiens à rappeler ma dette envers Ton Dees qui m'a accueilli au sein de son équipe en 1987 et m'a fait connaître les méthodes de recherche à l'aide de grands corpus informatisés, alors que je bénéficiais d'un congé de recherche sabbatique de l'Université de Montréal. Nous avons esquissé peu de temps avant son décès prématuré les plans d'un programme de cartographie comparée des données dialectales médiévales et modernes dans l'esprit de cette présentation. J'étends mes remerciements à tous les membres de l'équipe qui m'ont entouré et conseillé à Amsterdam et plus particulièrement Pieter van Reenen, Thera de Jong et Marcel Dekker. La recherche présentée ici est subventionnée en partie par le Conseil de Recherches en Sciences Humaines du Canada (*Les voies du français*, projet GTRC 412-2004-1002, sous la direction de Mme France Martineau). Les outils et les bases de données utilisées ici ont aussi été subventionnés en partie, depuis de nombreuses années, par le Conseil de recherches en sciences humaines du Canada et par le Fonds pour la formation des chercheurs et l'aide à la recherche du Québec — qui ont permis en particulier la saisie informatique des données cartographiques de l'ALF pour le domaine d'oïl, à laquelle ont participé activement Isabelle Beaulieu, Sophie Daoust, Jocelyn Guilbault, Louise Hétu, Caroline Mezzetta, Mylène Perrault, Phal Sok et Sandra Thibault: à tous, mes remerciements sincères.