

# L'interprétation dialectométrique des atlas 'scripturaux' d'Anthonij Dees

Autor(en): **Goebl, Hans / Smeka, Pavel**

Objektyp: **Article**

Zeitschrift: **Revue de linguistique romane**

Band (Jahr): **80 (2016)**

Heft 319-320

PDF erstellt am: **12.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-842211>

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern. Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden. Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

# L'interprétation dialectométrique des atlas 'scripturaux' d'Anthonij Dees

## 1. Prologue

Notre étude a comme objectif d'appliquer les méthodes de l'École Dialectométrique de Salzbourg (EDMS) aux données des atlas 'scripturaux' d'Anthonij Dees de 1980 et 1987. Nous avons déjà pu montrer par le passé, dès 1996, qu'une telle application est possible comme pour l'ALF ou tout autre atlas linguistique<sup>1</sup>. Notre idée est d'approfondir ainsi nos connaissances de ce que nous appelons, également depuis un certain temps, la 'gestion' ou l'aménagement linguistique de l'espace, que nous appréhendons ici pour le domaine d'oïl. Concrètement, nous souhaiterions montrer l'importance des calculs de localisation appliqués par Anthonij Dees, en 1983, à 222 textes littéraires du 12<sup>e</sup> et surtout du 13<sup>e</sup> siècle, pour une vision quantitative globale du domaine d'oïl médiéval<sup>2</sup>.

En dernière instance, ces nouvelles avancées sont dues à la jonction de deux traditions de recherche qui se sont développées, la plupart du temps, indépendamment l'une de l'autre: à savoir de la recherche scriptologique d'Amsterdam sous la conduite du regretté Anthonij Dees (1928-2001)<sup>3</sup>, et des travaux de l'EDMS<sup>4</sup>.

---

<sup>1</sup> Cf. *infra* et Goebel 1998, 2006, 2008a, ainsi que Goebel / Schiltz 2001.

<sup>2</sup> Le présent article constitue une refonte, corrigée et pourvue de nouveaux graphiques, d'une contribution scriptométrique déjà écrite en 2012 pour le compte de la revue *Medioevo romanzo* (MR) sous le titre «L'aménagement scripturaire du domaine d'oïl médiéval à la lumière des calculs de localisation d'Anthonij Dees effectués en 1983: une étude d'inspiration scriptométrique». Comme la publication de ce texte dans les volumes de MR s'est heurtée aux coûts de l'impression des 24 graphiques en couleurs, celui-ci avait été déposé à un endroit difficilement accessible du site internet officiel de MR. En accord avec les éditeurs de MR, nous proposons ici une version élargie de notre étude sous format papier.

<sup>3</sup> Cf. Dees 1984, 1985 et 1986, ainsi que les deux atlas scripturaux du même auteur (1980 et 1987). Voir aussi la nécrologie d'A. Dees rédigée par Lene Schøsler (*ici*, 2002), et le bilan global de ses recherches scriptologiques, dialectologiques et stemmatologiques publié par L. Schøsler et P. van Reenen en 2000.

<sup>4</sup> L'ouvrage-clé de l'EDMS est Goebel 1984. Voir, en outre, notre site internet avec une bibliographie, toujours tenue à jour, de nos contributions-DM: <[https://www.sbg.ac.at/rom/people/prof/goebl/dm\\_publi.htm](https://www.sbg.ac.at/rom/people/prof/goebl/dm_publi.htm)>.

À y regarder de près, ce rapprochement a connu deux moments cruciaux :

- Le premier se situe en 1996 quand, par l'aimable médiation de Piet van Reenen, les fichiers électroniques à l'aide desquels a été effectuée, en 1980, la publication de l'*Atlas des chartes du 13<sup>e</sup> siècle*, ont été transférés à Salzbourg et soumis immédiatement après à une dialectométrisation serrée par les soins de Guillaume Schiltz, alors boursier de la fondation de recherche autrichienne FWF<sup>5</sup> à Salzbourg<sup>6</sup>.
- Alors que le contact entre Amsterdam et Salzbourg, réalisé en 1996, est le résultat d'échanges et de contacts scientifiques de longue date et mérite, de ce fait, le qualificatif de 'normal', tout autre est le contexte dans lequel s'inscrit le second de ces deux mouvements de convergence. Car c'est ici que le hasard pur et simple a joué le rôle le plus important. Cette heureuse coïncidence n'a eu lieu ni à Amsterdam ni à Salzbourg, mais dans une petite localité située au cœur de la Forêt Noire.

## 2. La 'trouvaille de Lauterbad'

À l'occasion d'un atelier de recherche organisé, au mois de février 2006, par Achim Stein (Université de Stuttgart) et Pierre Kunstmann (Université d'Ottawa) à Lauterbad (Bade-Wurtemberg, Allemagne), dont le but était de traiter tous les aspects relatifs aux quelque 300 textes littéraires en ancien français qui constituent le *Nouveau Corpus d'Amsterdam*, Piet van Reenen, collaborateur de longue date d'Anthonij Dees, a fait circuler, parmi les convives d'un dîner amical en pleine Forêt Noire, deux ou trois liasses de papier<sup>7</sup>, issues d'une de ces imprimantes à aiguilles, aujourd'hui complètement démodées, et datées de novembre 1983. Ce faisant, Piet van Reenen a précisé que ces liasses lui avaient été remises par le bureau du Doyen de la Faculté des Lettres de l'Université Libre d'Amsterdam et que ce dernier était décidé à les jeter, ainsi que le reste du contenu de sept grands cartons abritant l'héritage scientifique d'A. Dees, découverts récemment dans les caves de l'Université, à moins que quelqu'un ne soit prêt à s'en occuper, en assumant évidemment le règlement des indispensables frais de transport.

Après un examen attentif du matériel présenté par Piet van Reenen, il s'est avéré qu'il s'agissait de calculs de localisation de plusieurs textes littéraires en ancien français effectués par A. Dees au Centre de calcul de l'Université Libre d'Amsterdam au mois de novembre 1983<sup>8</sup>.

<sup>5</sup> FWF – Fonds zur Förderung der wissenschaftlichen Forschung in Österreich.

<sup>6</sup> Cf. Goebel / Schiltz 2001, *passim*.

<sup>7</sup> Les dimensions en sont 38 sur 21 cm.

<sup>8</sup> Il semble bien que Dees ait commencé les travaux de localisation peu de temps après – ou même déjà avant – la publication de son atlas de 1980 : cf. à ce sujet la citation suivante qui se réfère à la fin de 1981 : « À l'heure actuelle (décembre 1981), quelque

Comme, à la vue de ces calculs, mon intérêt dialectométrique s'est éveillé sur le coup, j'ai<sup>9</sup> demandé à Piet van Reenen de me servir une seconde fois de médiateur et d'informer les responsables de l'Université Libre d'Amsterdam que j'étais entièrement disponible pour recevoir les sept cartons contenant le legs scientifique d'A. Dees. En effet, le transfert postal a été réalisé au cours de l'été de la même année (2006).

Après la réception des sept cartons annoncés et l'inspection de leur contenu, nous avons pu dresser le bilan suivant<sup>10</sup>:

- 222 liasses de calcul de 152 pages à structuration interne toujours identique. Ces liasses se réfèrent à autant de textes littéraires de l'ancien français, dont 182 se retrouvent également sur la liste des 200 textes de base de l'atlas scriptural de Dees de 1987.
- En outre, il y avait, dans les sept cartons en question, un nombre considérable de copies de textes littéraires dépouillés, le plus souvent recouverts d'annotations autographes d'A. Dees lui-même.

Par la suite, il a fallu analyser méticuleusement le contenu des liasses et surtout identifier les différentes étapes statistiques des calculs de localisation<sup>11</sup>. Ce faisant, nous avons constaté qu'il y avait, aux pages 146-150, des informations tant numériques que cartographiques qui semblaient se prêter à une dialectométrisation ultérieure.

Ces informations numériques, repérables dans chacune des 222 liasses, comprennent :

- pp. 146sq. : une liste de 87 valeurs de corrélation oscillant, grosso modo, entre 40,000 et >95,000 ; par la suite, nous appellerons le total de ces 87 valeurs 'vecteur de localisation' ;

---

150 textes ou fragments représentatifs de textes ont été examinés. Nous projetons de déterminer, dans les deux années à venir, la provenance de plusieurs centaines de textes ; il est probable qu'un deuxième atlas linguistique, basé sur les textes littéraires, pourra résulter de cette entreprise. » (Dees 1984, 110).

<sup>9</sup> L'emploi de la première personne du singulier renvoie à Hans Goebel seul alors que le pronom personnel *nous* se réfère aux deux auteurs de cet article.

<sup>10</sup> Dans toutes les recherches-DM relatives au legs scientifique d'A. Dees, j'ai été soutenu par mon collaborateur de recherche Pavel Smečka à qui j'adresse ici-même l'expression de ma sincère gratitude. C'est donc à juste titre qu'il figure comme co-auteur de cet article.

<sup>11</sup> Dans cette tâche nous avons reçu des informations et des conseils très précieux de la part d'Onno Huber (Université Libre d'Amsterdam) qui, à partir de 1980, faisait partie de l'équipe de Dees et dont le nom figure sur les frontispices des atlas de 1980 et 1987. Qu'il reçoive ici l'expression de notre profonde gratitude. Aux dires de M. Huber, une bonne partie des algorithmes de localisation utilisés par Dees avait déjà été mise au point par lui avant 1980.

- p. 148 : une mise en carte très rudimentaire des 87 centres scripturaux considérés par Dees en la matière, et
- pp. 149sq. : deux mises en carte, parfaitement identiques et graphiquement non moins rudimentaires, des scores de corrélation mentionnés ci-dessus. Elles ne concernent cependant pas les scores intégraux, mais les valeurs arrondies au nombre entier le plus proche, et dépourvues en plus de leurs décimales.

Ensuite, Pavel Smečka s'est attelé à la tâche de saisir électroniquement le total des 222 vecteurs de localisation repérés, alors qu'un autre collaborateur-DM de longue date, Slawomir Sobota, a entrepris d'élargir le réseau polygonal du domaine d'oïl relatif à l'atlas de 1980, de 85 à 87 entités territoriales<sup>12</sup>. Les deux entités (ou points) supplémentaires touchaient, d'un côté, l'Angleterre (point 87)<sup>13</sup> et, de l'autre, un centre scriptural en position plutôt centrale (point 86), baptisé par Dees « Aisne sud ».

### 3. Les calculs de localisation effectués par A. Dees en 1983 : buts et méthodes

#### 3.1. Les préliminaires dus à A. Dees

Retraçons, pour mettre au clair tant les antécédents que la finalité de notre démarche dialectométrique, le fonctionnement de la méthode déessienne de localisation telle qu'il l'a décrite à plusieurs reprises<sup>14</sup>. L'idée centrale en était de considérer les 85 (+2) vecteurs géographiques de l'atlas de 1980 – contenant, chacun, 268 des 282 attributs scripturaux figurant dans l'atlas de 1980 (= cartes no. 1-268; pp. 1-284) – comme autant d' 'étalons d'or' géographiques et de mesurer par la suite leurs similarités avec le vecteur d'attributs d'un des 222 textes littéraires sujets à localisation. Évidemment, ce vecteur d'attributs devait renseigner de manière quantitative sur la présence des 268 attributs scripturaux dans le texte littéraire en question. Il est non moins évident que la même procédure a dû être appliquée au reste des 222 textes de base.

<sup>12</sup> En 2006, nous disposions déjà d'un réseau polygonal du domaine d'oïl avec 85 points d'enquête qui avait été préparé par Guillaume Schiltz au cours de la dialectométrisation des données de l'atlas de Dees de 1980; cf. Goebel / Schiltz 2001. L'ajout de cellules polygonales à une mosaïque de tessellation déjà existante est, malheureusement, tout autre que trivial.

<sup>13</sup> Pour les données anglo-normandes qui ont dû être ajoutées après coup à la matrice de données de l'atlas (purement continental) de 1980, cf. de Jong 1988 et 1996, ainsi que Schøsler / van Reenen 2000, 39.

<sup>14</sup> Cf. les avant-propos des deux atlas scripturaux (1980 et 1987), ainsi que les contributions suivantes d'A. Dees: 1984, 1985 et 1986.

Rappelons que le chiffre de 268 (= nombre des attributs scripturaux considérés dans les calculs de Dees effectués en 1983) est légèrement inférieur au nombre des attributs scripturaux cartographiés dans l'atlas de 1980 (= 282), mais surtout de loin inférieur au 517 attributs de l'atlas de 1987.

Le résultat de chacune de ces 222 comparaisons, exécutées à l'aide d'un algorithme de corrélation<sup>15</sup> dont on trouve une description sommaire dans l'introduction de Dees 1987 (XIX-XXX), consiste en 87 scores de corrélation (situés théoriquement entre -100 et +100) qui se répartissent sur toute la surface du domaine d'oïl.

Dans la conception 'localisatrice' de Dees, la valeur maximale des 87 scores calculés équivaut à l'entité spatiale où, selon toute probabilité, le manuscrit en question a été rédigé. Pour cette raison, les valeurs maximales calculées figurent toujours, dans l'atlas de 1987, parmi les éléments descriptifs des 200 textes littéraires sur lesquels il repose (voir à ce sujet Dees 1987, 519-533)<sup>16</sup>.

Pour le grand public, c.-à-d. pour les lecteurs des atlas de 1980 et 1987, il existe cependant un grand inconvénient: sur les cartes de ces deux atlas (282 cartes pour l'atlas de 1980; 517 cartes pour l'atlas de 1987), nul ne peut percevoir ni 85 ni 87 entités territoriales. Les planches de l'atlas de 1980 ne montrent que 28 'centres scripturaux' alors que celles de l'atlas de 1987 n'en présentent que 29. Pourtant, dans les «préliminaires» de ce dernier (xvissqq.), il est question de 87 «points géographiques».

En fait, les effectifs *majeurs* (85 à la place de 28 entités territoriales, et 87 au lieu de 29) existent *uniquement* dans les bases de données électroniques 'sous-jacentes' dont Dees et nous-mêmes nous sommes servis.

Le but primordial des calculs de localisation de Dees était donc de 'calibrer' géographiquement un nombre relativement grand de textes littéraires provenant de tous les coins du domaine d'oïl, et de conférer ainsi un certain degré de sûreté diatopique, empruntée aux «témoins primaires<sup>17</sup>» (= les 3 300

<sup>15</sup> Aux dires d'Onno Huber, il s'agissait, au début des années 80, non pas d'un seul algorithme de corrélation, mais de tout un jeu de tels algorithmes dont les modalités ont été souvent définies par Dees d'une manière très spontanée, pratiquement en face des grandes machines calculatrices ('supercalculateurs', 'main frame') alors en usage.

<sup>16</sup> Signalons que la base de données du «Nouveau Corpus d'Amsterdam» (NCA) répertorie également ces valeurs maximales; voir à ce sujet: <<http://julienas.philosophie.uni-stuttgart.de/nca/version2/download/amslitbib.html>>. Il serait d'ailleurs fort utile d'y présenter aussi – dans la mesure où les textes localisés par Dees en 1983 et ceux du NCA se recoupent – les profils de localisation respectifs dans toute leur plénitude heuristique telle que celle-ci ressort des cartes 1-8 de cet article.

<sup>17</sup> Cf. Dees 1987, XIII.

chartes de son atlas de 1980), à ce qu'il appelait les « témoins dérivés » ou carrément « suspects<sup>18</sup> » (= les textes littéraires prévus pour son atlas de 1987).

Il semble bien que Dees ait utilisé les résultats de ses calculs de localisation pour la définition du corpus de base de son atlas de 1987 qui repose – comme nous l'avons dit – sur 200 textes littéraires. Il est curieux de voir, d'un côté, que Dees indique, dans son atlas de 1987 (519-533), pour chacun des 200 textes de base un 'coefficient' de localisation, et de constater, de l'autre, que pour 18 des 200 textes en question, il n'y a pas d'équivalent dans les 222 liasses détectées dans le sous-sol de l'Université Libre d'Amsterdam.

Deux questions s'imposent :

- Peut-être la campagne de localisation de 1983 avait-elle été reprise (ou en partie corrigée) au cours des travaux de préparation pour l'atlas publié en 1987<sup>19</sup> ?
- Étant donné que nous disposons de 222 calculs de localisation dont seulement 182 ont laissé des traces dans l'atlas de 1987, à quoi servaient (et où se sont retrouvés) les 40 (= 18 + 22) calculs de localisation apparemment 'superflus' ?

Après la date fatidique de 1983, Dees avait donc opéré une sélection dont il a indiqué les critères dans la préface (xvii) de son atlas de 1987<sup>20</sup>.

Ajoutons encore une précision indispensable pour une bonne compréhension de la procédure localisatrice à laquelle Dees avait soumis les 200 textes de son atlas de 1987. Il en avait besoin pour assigner, à chacun de ces 200 textes littéraires, tous (géo)linguistiquement très composites voire hétéroclites, l'endroit statistiquement « le plus sûr » parmi les 87 « points » du réseau de l'atlas en question. De cette manière, il pouvait déterminer, avec une précision sensiblement accrue, la répartition géographique des 517 attributs scripturaux analysés dans les textes sélectionnés pour son atlas de 1987. C'est ainsi qu'il

<sup>18</sup> Cf. Dees 1987, XIV.

<sup>19</sup> À la fin de la préface de son atlas de 1987 (XXXI), Dees mentionne les années 1980-1984 comme période décisive dans la préparation de cet atlas.

<sup>20</sup> Voir aussi les précisions signalées chez Schøsler / van Reenen 2000, 40-43. Pour la constitution du corpus de base de son atlas de 1987, Dees avait admis uniquement des textes littéraires dont le vecteur de localisation comprenait un score maximal situé au-dessus de 70. Ce faisant, il écartait un total de 35 textes, préalablement localisés, dont les vecteurs de localisation détenaient des scores maximaux au-dessous de 70. Ceci démontre clairement que Dees valorisait, dans sa conception localisatrice, exclusivement le score maximal des 87 coefficients de localisation calculés pour chacun des textes littéraires examinés, et non pas, comme c'est le cas ici, le bilan global de la mise en carte du contenu intégral des vecteurs de localisation relevés. Encore un détail : d'une comparaison des matériaux qui figurent dans la 'trouvaille de Lauterbad' avec ceux du NCA, il résulte que des 222 textes de Lauterbad (= Dees 1983), 45 n'ont pas de contrepartie dans les 300 textes du NCA.

a réussi à transférer la fiabilité diatopique des chartes (« témoins primaires ») au traitement géolinguistique des attributs scripturaux des textes littéraires (« témoins dérivés » ou même « suspects »)<sup>21</sup>.

### 3.2. Les prolongements scriptométriques effectués à Salzbourg

Voir la figure 1.

Jetons, pour approfondir nos connaissances de la démarche de Dees et de celle de l'EDMS, un regard sur la figure 1 (cf. *infra*). La partie supérieure en montre la localisation d'un texte littéraire isolé. Le résultat est un vecteur de localisation avec 87 valeurs dont les scores s'inscrivent dans un espace numérique situé grosso modo entre -40 et +95 (cf. *supra*, 2.).

Alors que l'intérêt de Dees s'épuisait dans la constatation des scores maximaux de ces vecteurs de localisation, l'EDMS visait d'entrée de jeu le total des 87 scores de localisation en question pour en tirer un profil choroplèthe finement articulé tout en le considérant comme 'jauge' de l' 'oïlité' diatopique du texte littéraire respectif.

Évidemment, il fallait rendre visibles tous ces gradients spatiaux en leur conférant une visualisation adéquate selon les modalités de l'EDMS<sup>22</sup> : voir la partie inférieure de la figure 1 et les cartes 1-8.

La visualisation pure et simple des 222 gradients de localisation est loin d'en épuiser toutes les possibilités taxométriques. Le pas décisif vers une analyse taxométrique plus enveloppante consiste dans la réunion des 222 vecteurs de localisation calculés par Dees en 1983, sous la forme d'une *matrice de données factice* (de nature quantitative). Cette opération repose sur l'hypothèse qu'il est possible de capter ainsi l'aménagement scriptural du domaine d'oïl médiéval du côté littéraire. Ce faisant, nous formulons, en outre, l'hypothèse qu'un dépouillement ultérieur de cette matrice de données factice suivant les normes de l'EDMS puisse aboutir à des résultats classificatoires comparables à ceux que nous avons déjà obtenus à l'aide de la matrice de données de base de l'atlas scriptural de Dees de 1980<sup>23</sup>.

<sup>21</sup> Bien que trente ans se soient écoulés depuis, il n'est pas trop tard pour rendre hommage à l'inventeur de cette procédure quantitative carrément géniale.

<sup>22</sup> Précisons que Dees a publié plusieurs des « cartes numériques de localisation », produites au cours de ses calculs de localisation : cf. Dees 1980, 370 ; 1984, 117-118 ; 1985, 115 et 1986, 517. Vu leur mauvaise qualité cartographique, le rendement heuristique en est très limité.

<sup>23</sup> Voir nos contributions de 1998, 2001, 2006, 2008a et 2011, ainsi que Goebel / Schiltz 2001.





Que cette hypothèse mérite en effet son nom s'explique par le fait suivant : un parallélisme complet entre les dialectométrisations des deux atlas de Dees (de 1980 et 1987) eût rendu nécessaire la disponibilité parallèle des deux matrices de données (réelles) de base. Or, en 1997, il s'est avéré que la matrice de données de base de l'atlas de 1987 avait été égarée, entre cette date et 1997, dans les méandres du Centre de calcul de l'Université Libre d'Amsterdam<sup>24</sup>. Vu que l'atlas de 1987 ne contient, dans sa forme imprimée, que le réseau au maillage relâché de 29 « points » (ou : unités spatiales), il est impossible de reconstituer le contenu de la matrice de données originale qui, elle, était basée sur 87 unités spatiales. Il s'agit donc d'une perte irréparable.

Évidemment, cette déperdition très regrettable rehausse considérablement la valeur et l'importance de la 'trouvaille de Lauterbad' bien que, du point de vue algorithmique, celle-ci n'ait pas exactement le même statut statistique que la matrice de données de base égarée. Or, les résultats de l'analyse-DM effectuée à l'aide de la matrice de données factice<sup>25</sup> confirment pleinement la validité de notre hypothèse de départ, à savoir qu'il est possible de capter l'aménagement scriptural (littéraire) du domaine d'oïl médiéval au travers de celle-ci.

La figure 1 nous montre, en outre, que le dépouillement-DM ultérieur de la matrice de données factice passe par l'application d'un indice de similarité (ou de corrélation) adéquat pour la saisie des similarités (quantitatives) entre les 87 centres scripturaux<sup>26</sup>, pour aboutir à l'établissement d'une matrice de similarité (carrée) qui, elle, servira de point de départ pour l'application de toutes les méthodes de dépouillement, *numériques* et – surtout – *visuelles*, développées à Salzbourg<sup>27</sup>.

Dans le paragraphe 5, nous en présenterons quelques échantillons tout en les comparant régulièrement avec les résultats analogues tirés de l'atlas de 1980 et aussi de l'ALF. Il en dérivera une optique comparative captivante qui embrasse un laps de temps de six siècles.

Remarque technique : comme le logiciel salzbourgeois VDM (« Visual DialectoMetry ») ne tolère que des valeurs entières et positives, nous avons dû appliquer, aux scores originaux calculés par Dees en 1983, une transformation

<sup>24</sup> Cf. à ce sujet, Schøsler 2002, 617.

<sup>25</sup> Cette matrice de données factice comprend 87 centres scripturaux et 222 « attributs » : voir la partie inférieure de la figure 1.

<sup>26</sup> Il s'agit du « coefficient de corrélation de Bravais et Pearson » [r(BP)] : pour une description détaillée, cf. Sneath / Sokal 1973, 138.

<sup>27</sup> Cf. à cela, Goebel 1984 (*passim*) ou Goebel 2008a (comme modèle méthodique).

mathématique relativement simple. Ayant constaté que le score *minimal* repérable dans le total des calculs de Dees accomplis en 1983 est de -43,255, et que le score *maximal* se situe vers >97, nous avons augmenté tous les scores calculés par Dees du montant de >43 et supprimé les valeurs décimales, si bien que la variabilité numérique traitée par le logiciel VDM s'inscrit entre 0 et >140.

#### 4. Présentation à titre d'exemple de huit profils de localisation

Voir les cartes 1-8.

Les 222 vecteurs de localisation calculés par Dees représentent, après leur mise en carte, autant de profils choroplèthes finement articulés où les scores de corrélation se fondent, sans exception aucune, dans un gradient spatial aux contours finement ondulés, et agencé très souvent de façon auréolaire. La structuration régulière de ces profils est absolument épatante. Elle suggère que l'activité communicative des scribes médiévaux – quelles que soient leur origine personnelle et/ou leur patrie linguistique – était sujette aux mêmes régularités (voire lois) et contraintes spatiales que l'activité langagière des dialectophones du temps de l'ALF, telle que celle-ci a été révélée par nos analyses-DM antérieures<sup>28</sup>. Le fait que les premiers s'exprimaient dans une modalité et conception scripturales et les deuxièmes à l'oral ne met pas en cause leur inscription dans les mêmes contraintes de l'espace.

Le choix des huit échantillons présentés est uniquement dicté par des besoins géographiques et n'a d'autre but que de donner aux lecteurs une impression cavalière des grands paysages scripturaux du domaine d'oïl médiéval:

carte 1: *Sud-Ouest*: «Der sogenante poitevinische Pseudo-Turpin», éd. Th. Auracher, in: *ZrP* 1 (1877), 259-336, ici p. 262,1-294,22 (NCA 216: date ms.: 1<sup>er</sup> q. 13<sup>e</sup> s.)

carte 2: *Angleterre*: «Bestiaire de Guillaume le Clerc», éd. R. Reinsch, Wiesbaden 1892 (NCA 36: date composition 1211 / date ms.: ca 1250)

carte 3: *Centre*: «Ms. V du Perceval de Chrétien de Troyes», transcription R. L. H. Rops (NCA 214: date composition ca 1180 / date ms.: 1<sup>er</sup> t. 14<sup>e</sup> s.)

carte 4: *Picardie*: «Renart le Nouvel par Jacquemart Glielee», v. 1-3026, éd. H. Roussel, SATF, Paris 1961 (NCA 191: date composition ca 1290 / date ms.: 1295)

carte 5: *Wallonie*: «Médicinaire liégeois du XIII<sup>e</sup> siècle», éd. J. Haust, Bruxelles 1941 (NCA 65: date composition 2<sup>e</sup> m. 13<sup>e</sup> s. / date ms.: 2<sup>e</sup> m. 13<sup>e</sup> s.)

<sup>28</sup> Cf. Goebel 2002 et 2003.

carte 6: *Lorraine*: «Dialogus anime conquerentis et rationis consolantis» (éd. F. Bonnardot, in: *R* 5 (1876) 269-332 (NCA 30: date composition *ca* 1200 / date ms.: *ca* 1200)

carte 7: *Franche-Comté*: «Des Poignes d'enfer, les deux chevaliers, enseignement moral», éd. P. Meyer, in: *R* 6 (1877) 1-46 (NCA 40: date ms.: 1<sup>er</sup> q. 14<sup>e</sup> s.)

carte 8: *Bourbonnais*: «Roman de la Rose de Guillaume de Lorris», éd. F. Lecoy, CFMA, Paris 1965 (NCA 245: date composition *ca* 1230 / date ms.: *ca* 1285)

Évidemment, tous les textes mentionnés ci-dessus figurent dans le corpus de l'atlas de 1987 (519-533).

Remarque technique: les valeurs maximales repérables dans les légendes des cartes 1-8 ne correspondent pas à ce que l'on peut trouver chez Dees (1987, 519-533), à cause de leur transformation numérique préalable par l'addition de la valeur constante de +43<sup>29</sup>.

Sur les cartes 1-8, la visualisation (ou mise en carte) des scores de localisation obéit aux principes de l'algorithme de visualisation (ou d'intervallisation) MINMWMAX 6-tuple (avec six paliers chromatiques) qui, au sein de l'EDMS, représente le procédé de visualisation standard<sup>30</sup>.

Les couleurs *chaudes* (rouge, orange, jaune) reflètent les scores situés *au-dessus* de la moyenne arithmétique (tirée à partir des 87 valeurs de localisation), alors que les couleurs *froides* (bleu foncé, bleu moyen, vert) correspondent à ceux qui se trouvent *au-dessous* de cette dernière. Les seuils numériques intermédiaires résultent de la tripartition des écarts numériques repérables – d'une part – entre le *minimum* et la *moyenne arithmétique* et – de l'autre – entre le *maximum* et cette dernière.

L'agencement général des six couleurs utilisées suit la gamme chromatique de l'arc-en-ciel (équivalente à celle du spectre solaire).

## 5. Cinq comparaisons dialectométriques à trois volets (Dees 1983 – Dees 1980 – ALF)

Nous allons présenter ci-dessous six comparaisons dialecto- ou scriptométriques entre trois stocks de données:

- les calculs de localisation de 1983 (voir les cartes de *gauche*),
- la dialectométrisation des données de l'atlas de 1980 (voir les cartes du *milieu*), et

<sup>29</sup> Voir aussi notre remarque faite à la fin du paragraphe 3.

<sup>30</sup> Voir à ce sujet, Goebel 1981, 361-363 (explication en français) et 1984 I, 93-94 (présentation en allemand).

- la dialectométrisation des données de la partie oïlique du réseau de l'ALF (voir les cartes de *droite*) dont le maillage a été préalablement réduit à la densité des réseaux utilisés par Dees en 1980 et 1983 (cf. *infra* 5.1).

Il en résulte une étude comparatiste – heuristiquement finement équilibrée et linguistiquement très intéressante pour ne pas dire fascinante – dont l'ampleur diachronique embrasse quelque 600 ans. Le seul bémol reste que les cartes de l'ALF reflètent une véritable synchronie (années 1897-1901), alors que celles de l'*Atlas des chartes* couvrent essentiellement les cinquante années 1250-1300 et celles de l'*Atlas des textes littéraires* une tranche chronologique plus large de la fin du 12<sup>e</sup> au début du 14<sup>e</sup> siècle.

Les cinq niveaux comparatifs se réfèrent aux méthodes-DM suivantes: cartes de similarité, cartes interponctuelles, cartes à paramètres (concernant les maxima et les CAF), analyse dendrographique et cartes à corrélations.

### 5.1. Comparaison dialectométrique à trois volets (Dees 1983 – Dees 1980 – ALF) A: vingt-et-une cartes de similarité

Voir les cartes 9-29.

Les cartes 9-29 fournissent une comparaison directe entre la dialectométrisation des deux stocks de données de Dees (1983 et 1980) et des données 'diluées' de l'ALF. Les ressemblances iconiques qui en ressortent sont vraiment remarquables. Inutile de dire que nous avons appliqué, pour toutes les comparaisons à trois volets, les mêmes conventions *statistiques* (c.-à-d. pour le calcul des similarités entre le point de référence *j* et les points comparés *k*) et *cartographiques* (c.-à-d. pour la visualisation par voie algorithmique):

- calcul de la similarité (par points d'atlas appariés [*j* et *k*]):  
données de Dees (1983 et 1980):  
coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]^{31}$   
données de l'ALF:  
Indice Pondéré d'Identité (avec le poids 1)  $[IPI(1)_{jk}]^{32}$
- visualisation:  
algorithme d'intervallisation MINMWMAX 6-tuple: voir supra, fin du paragraphe 4.

La densité des deux réseaux-Dees n'est pas la même: alors que l'atlas de Dees de 1980 comprend, sur le continent, 85 centres scripturaux, le réseau

<sup>31</sup> Pour une explication, cf. Sneath / Sokal 1973, 138.

<sup>32</sup> Pour une présentation détaillée de l' $IPI(x)_{jk}$ , voir Goebel 1987, *passim*.

utilisé par Dees en 1983 (et aussi dans son atlas de 1987) en contient 86 (avec l'ajout du point 86, Aisne sud) auxquels s'ajoute le centre scriptural (détaché) de l'Angleterre (point 87).

Une comparaison supplémentaire avec les données de l'ALF s'impose. À cet effet, nous avons 'dilué' le réseau original de l'ALF dans sa moitié septentrionale en correspondance avec l'étendue traditionnelle du domaine d'oïl. Il en résulte un réseau réduit de 85 points-ALF dont les distances géographiques moyennes équivalent grosso modo à celles du réseau de Dees 1980.

La quantité des données-ALF (c.-à-d. des 'cartes de travail', CT) utilisées pour les calculs-DM à l'échelle d'oïl, n'est pas exactement la même que pour les calculs-DM effectués à l'échelle originale de l'ALF<sup>33</sup>: ceci s'explique par l'élagage général du maillage de l'ALF et l'amputation de la moitié sud du réseau de l'ALF: voir le tableau synoptique suivant et aussi la figure 2.

	Réseau-ALF intégral	Réseau-ALF élagué
Σ points d'enquête	638+3 <sup>34</sup>	85
Σ cartes-ALF originales dépouillées	626	587
Σ cartes de travail (CT)	1 681	1 279
Σ taxats / aires taxatoires <sup>35</sup> (AT)	18 047	6 302
Polynymie interne des CT	de 2 à 90 taxats / AT par CT	de 2 à 23 taxats / AT par CT

Les cartes de travail des deux corpus embrassent toutes les catégories linguistiques (phonétique, morpho-syntaxe et lexique) prises en considération au moment de la dialectométrisation intégrale de l'ALF. La figure 2 visualise la distribution quantitative des 1 279 cartes de travail du corpus-ALF élagué dont la fragmentation interne (polynymie<sup>36</sup>) varie entre 2 et 23 taxats (ou aires taxatoires).

<sup>33</sup> Cf. à ce sujet, Goebel 2002 et 2003.

<sup>34</sup> Rappelons que le nombre des points-ALF est de 638 et que nous y avons ajouté, au cours de la dialectométrisation de l'ALF, trois points artificiels: pour plus de détails, cf. Goebel 2002 et 2003.

<sup>35</sup> Par «aire taxatoire», l'on entend, au sein de l'EDMS, l'étendue occupée par un type linguistique («taxat») sur une carte muette de l'atlas linguistique examiné. Comme nul taxat ne peut se passer d'une aire de distribution, le nombre des AT est toujours égal à celui des taxats à proprement parler.

<sup>36</sup> Pour l'importance linguistique du phénomène et du concept de la polynymie en matière de DM, cf. Goebel 2014.

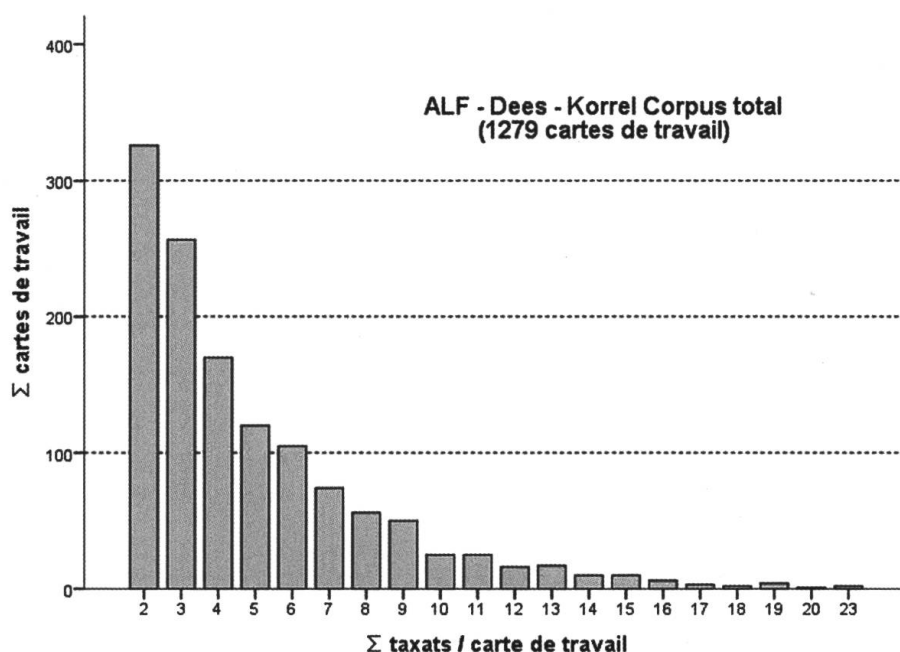


Figure 2: Histogramme – avec 20 barres verticales de hauteurs inégales – montrant, pour les 1 279 cartes de travail (CT) et les 6 302 taxats (aires taxatoires) figurant dans le corpus-ALF élagué, la relation entre la polynymie interne des CT (en abscisse) et le nombre absolu des CT respectifs (en coordonnée).

Exemple de lecture: il y a 326 CT *di*-nymes (= CT avec *deux* taxats) et 50 CT *ennéa*-nymes (= CT avec *neuf* taxats).

Le bilan de la considération synoptique des sept triplets cartographiques est très net: les structures de profondeur des trois stocks de données se ressemblent beaucoup, et ceci malgré la grande différence entre d'une part le caractère scriptural des données de Dees et l'oralité des données de l'ALF, et d'autre part le décalage temporel entre les deux 'points' de repère diachroniques (1200/1300 et 1900). Nous sommes donc en droit de dire qu'à la lumière de ces parallélismes structuraux, les bases sous-jacentes de la fragmentation diatopique du domaine d'oïl semblent répondre à des principes identiques au 13<sup>e</sup> siècle et en 1900. Ceci signifie qu'il faut repousser la question de la genèse de la fragmentation dialectale du domaine d'oïl à une période située bien avant la date fatidique de *ca* 1200<sup>37</sup>.

<sup>37</sup> Voir les questions soulevées à ce sujet dans les ouvrages de J. Wüest (1979) et L. Remacle (1992). Notre constat rejoint pleinement les résultats de Carles (2016 et 2017) selon laquelle la stratification spatiale des romanismes dans les textes latins avant 1100 est concomitante à la genèse des langues galloromanes.

En règle générale, l'on peut constater que la ressemblance réciproque entre les profils choroplèthes situés à gauche (Dees 1983) et au milieu (Dees 1980) est plus grande que celle qui se dessine entre ces derniers et le profil choroplèthe situé à droite (relatif à l'ALF). En ligne générale, les données littéraires sont plus homogènes dans la distribution spatiale que les données documentaires, ce qui devrait s'expliquer par l'effet des copies, propres aux textes non-documentaires et induisant un caractère plus composite et plus dilué (notre 'effet tache d'huile'). Mais la convergence entre les deux ensembles est notable, surtout dans la synthèse générale (cf. *infra*, cartes 43 et 44). Le décalage par rapport aux données de l'ALF n'est en revanche pas surprenant, étant donné la distance chronologique d'un demi millénaire. L'on relève notamment les effets d'un brassage linguistique des dialectes parlés au centre du domaine d'oïl, survenu entre les 13<sup>e</sup> et 19<sup>e</sup> siècles<sup>38</sup>. Nous en reparlerons dans les paragraphes suivants.

### 5.2. Comparaison dialectométrique à trois volets (Dees 1983 – Dees 1980 – ALF) B: trois cartes interponctuelles

Voir les cartes 30-32.

L'analyse interponctuelle fournit des cartes dites 'isarithmiques' où apparaissent, le long des côtés de polygone, des phénomènes de cloisonnement qui rappellent les faisceaux d'isoglosses de la géographie linguistique traditionnelle. Les effets de cloisonnement tiennent à deux raisons : à l'accroissement de l'épaisseur et au coloriage contrasté (en bleu foncé) des côtés de polygone répertoriés dans l'intervalle 6.

La syntaxe iconique de la carte 30 (relative au réseau de Dees 1983) repose sur 229 côtés de polygone, alors que celle des cartes 31 et 32 (relatives au réseau de Dees 1980 et à celui de l'ALF élagué) en contient 225. La différence s'explique par la présence supplémentaire de deux polygones dans le réseau de 1983.

Par rapport à l'algorithme standard de visualisation MINMWMAX, l'algorithme utilisé pour la confection des cartes 30-32 – MEDMW (6-tuple) – rehausse l'impact iconique des intervalles 5 (en bleu moyen) et 6 (en bleu foncé) afin que la vertu séparatrice des côtés de polygone concernés ressorte mieux<sup>39</sup>.

<sup>38</sup> Voir aussi Goebel 2008a, 17-19.

<sup>39</sup> Comme l'analyse interponctuelle sert à la saisie quantitative de *différences* observables entre deux points d'atlas contigus, elle rend nécessaire l'emploi d'indices de *distance*, et non plus de *similarité*. Sur les cartes 30-31, il s'agit, en l'occurrence, de l'indice  $DEM_{jk}$  (Distance Euclidienne Moyenne) et, sur la carte 32, de l'indice  $IPD(1)_{jk}$  (Indice Pondéré de Distance, avec le poids 1). Voir, pour une définition de



De la considération comparative globale des cartes 30-32, il ressort une divergence remarquable : elle concerne la structuration isoglottique du centre du domaine d'oïl. Alors que celui-ci est caractérisé par un sillonnement interponctuel relativement intense sur les cartes 30 et 31, il revêt un aspect complètement différent sur la carte 32 qui reflète la situation interponctuelle à la fin du 19<sup>e</sup> siècle (ALF). L'on y perçoit une zone largement égalisée aux contours bien définis vers le Sud-Ouest et le Cotentin, et surtout vers le Nord (Picardie, Artois), l'Est (Lorraine) et le Sud-Est (Franche-Comté).

La raison d'être de cet état des choses est facilement repérable : il s'agit des effets du rayonnement pluriséculaire du type linguistique d'oïl à partir de l'Île-de-France, survenu entre les 13<sup>e</sup> et 19<sup>e</sup> siècles.

### 5.3. Comparaison dialectométrique à trois volets (Dees 1983 – Dees 1980 – ALF) C: dix cartes à paramètres, relatives aux valeurs maximales (*maxima*) et aux coefficients d'asymétrie de Fisher (CAF)

#### 5.3.1. La synopse des valeurs maximales

Voir les cartes 33-35.

Comme le démontrent les histogrammes figurant dans la partie inférieure des cartes 9-29, toute carte de similarité repose, en dernière analyse, sur une distribution de fréquence (dite aussi «de similarité»). Or, toute distribution de fréquence comprend un certain nombre de valeurs caractéristiques (ou «paramètres»), telles la *moyenne arithmétique*, l'*écart-type*, la *médiane*, les valeurs extrêmes (*minimum* et *maximum*) et aussi quelques valeurs supplémentaires qui en saisissent la plus ou moins grande *symétrie* (ou *asymétrie*). Un des acquis majeurs de l'EDMS est d'avoir découvert l'importance de la présentation synoptique de ces paramètres pour les fins de la géolinguistique.

Dans cette section, il est question de la synopse des *valeurs maximales* des distributions de similarité; elle sert à mettre en évidence un concept que la dialectologie traditionnelle de type qualitatif avait souvent évoqué sans d'ailleurs avoir réussi à le traiter avec précision : il s'agit des 'noyaux dialectaux'<sup>40</sup>. L'idée générale en était que n'importe quel paysage dialectal pouvait être caractérisé par la concomitance alternée de *noyaux* dialectaux et de *zones de*

ces indices, Goebel 2008a, 29-30 (pour la  $DEM_{jk}$ ), et 1987, 70-79 (pour l' $IPD(1)_{jk}$ ). Précisons qu'entre les indices pondérés  $IPD(x)_{jk}$  et  $IPI(x)_{jk}$ , il existe la relation (de complémentarité) suivante:  $IPD(x)_{jk} + IPI(x)_{jk} = 100$  (ou: 1).

<sup>40</sup> Voir aussi Goebel 1981, 382-388 (explication en français) et 1984 I, 140-143 (explication en allemand).

*transition*, ces dernières se creusant harmonieusement entre les noyaux dialectaux à l'instar des vallées entre les montagnes environnantes.

À l'intérieur d'un réseau géolinguistique, la taille de la similarité<sup>41</sup> entre deux lococlectes contigus peut varier considérablement. Si donc plusieurs lococlectes voisins se partagent des scores de similarité réciproque très grands, c.-à-d. situés au-dessus de la moyenne arithmétique de tous les scores de similarités calculés, il en résulte, sur la mise en carte respective, une implantation *ponctuelle* colorée en rouge, orange ou jaune qui, elle, est entourée de configurations *linéaires* marquées en vert, bleu moyen ou bleu foncé. Évidemment, ces dernières reflètent des scores de similarité répertoriés au-dessous de la moyenne arithmétique en question. Or, c'est exactement cet effet cadencé qui peut être constaté sur les cartes 33-35.

Sur la carte 33, ce sont – en sens horaire – la Picardie, la Lorraine, la Franche-Comté et le Poitou qui se particularisent comme 'noyaux'. Tel est le cas aussi sur la carte 34, relative aux documents non littéraires, bien qu'avec certains décalages par rapport à la carte 33.

Quant à la carte 35, relative à la période de l'ALF, l'implantation des noyaux dialectaux qui s'y dégagent est quelque peu différente de celle des cartes 33 et 34. À l'instar de la carte 31, le centre du domaine d'oïl y ressort avec beaucoup de netteté et aussi de cohérence (voir les plages polygonales en rouge et orange, classées dans les intervalles 6 et 5), alors que la Picardie (voir les trois polygones en rouge et orange), la Wallonie (voir les deux polygones en jaune) et la Franche-Comté (*idem*) occupent la position de noyaux périphériques.

### 5.3.2. La synopse des « coefficients d'asymétrie de Fisher » (CAF)

Voir les cartes 36-42.

Parmi les paramètres statistiques susceptibles d'être mis en synopse, le « coefficient d'asymétrie de Fisher » (CAF) occupe une place de choix<sup>42</sup>. C'est un indice qui sert à la saisie quantitative de la symétrie d'une distribution de similarité. Or, le degré de similarité (quantitative) d'une distribution de similarité dispose d'une grande importance linguistique.

Plus la silhouette d'une distribution de similarité penche vers la gauche, plus grand est le nombre des scores de similarité situés *au-dessous* de la

<sup>41</sup> L'indice de similarité utilisé ( $SEM_{jk}$ ) est le complément de la « Distance Euclidienne Moyenne » ( $DEM_{jk}$ ):  $100 - DEM_{jk}$ .

<sup>42</sup> Pour une présentation circonstanciée du CAF, cf. Goebel 1981, 394-401 (en français) et 1984 I, 150-153 (en allemand).

moyenne arithmétique. L'interprétation *linguistique* de cette asymétrie est que la portée *communicative* du locolecte en question est *mauvaise*.

La situation inverse se présente quand la majorité des scores de similarité s'amoncelle entre la moyenne arithmétique et le maximum de la distribution de similarité respective : dans ce cas, la silhouette de la courbe en question se penchera vers la droite et, de ce fait, la portée communicative du locolecte examiné s'améliorera considérablement.

Il en résulte que le CAF constitue un excellent moyen-DM pour la saisie (par voie quantitative) d'une des propriétés centrales de n'importe quel réseau interactif, à savoir le degré d'interaction et d'échange avec lequel chaque membre (ou élément) du réseau en question participe aux flux et courants de communication qui, inévitablement, circulent dans le maillage du réseau analysé.

Dans cette perspective, l'on peut distinguer, parmi les éléments d'un maillage de points d'enquête<sup>43</sup>, les 'hyperactifs' d'un côté, et les 'abstentionnistes' de l'autre. Évidemment, la relation (en %) qui existe entre ces deux types d'interaction varie considérablement d'un point d'enquête à l'autre. Il est donc indiqué de saisir la palette de ces fluctuations et de l'analyser du point de vue linguistique. Or, de tels phénomènes correspondent à ce qu'il est convenu d'appeler 'brassage' (ou 'compromis') linguistique en français, et 'Sprachausgleich' en allemand.

Au regard des 6 302 aires taxatoires de notre corpus-ALF élagué, toutes géographiquement très multiformes<sup>44</sup>, le phénomène du brassage linguistique peut être expliqué de la manière suivante : l'intensité en sera considérable dans les zones du réseau où la majorité des points d'enquête respectifs dispose d'aires taxatoires très grandes ou en voie d'élargissement, alors que l'inverse est vrai pour les zones où circulent des aires taxatoires moins dynamiques dont l'étendue est plutôt restreinte ou en voie de rétrécissement.

Le CAF – qui constitue une des pièces fortes de l'EDMS – permet donc de saisir une des propriétés les plus saillantes (et aussi déroutantes) de n'importe quel atlas linguistique, à savoir l' 'enchevêtrement' (ou le 'tuilage') particulier<sup>45</sup> des aires taxatoires, et de créer ainsi des profils choroplèthes qui sont très utiles du point de vue surtout *diachronique*.

<sup>43</sup> N'oublions pas que tout maillage de points d'enquête est une manifestation idéale d'interactions spatiales à plus ou moins longue distance.

<sup>44</sup> Cette multiformité s'exprime dans trois dimensions : taille, cohérence et positionnement des aires taxatoires au sein du réseau de points examiné.

<sup>45</sup> Le caractère instable, voire primesautier, de l' 'enchevêtrement particulier' typique des aires taxatoires n'a pu être saisi, accepté et compris que très lentement par les géolinguistes qui, eux, se sont longtemps obstinés à en postuler la coïncidence – ou au moins convergence – spatiale aussi parfaite que possible.

Pour mieux faire ressortir le message linguistique des profils choroplèthes des cartes 36-42, nous ne les avons visualisées qu'à l'aide de deux paliers chromatiques<sup>46</sup>. C'est ainsi que les plages polygonales en rouge renvoient à des zones dont le comportement *scriptural* (voir les cartes de gauche et du milieu) et *dialectal* (voir les cartes de droite) est proche de celui des 'abstentionnistes' évoqués tout à l'heure, alors que les polygones marqués en bleu symbolisent des régions où se démènent les 'hyperactifs' et où, de ce fait, le brassage linguistique est intense.

Pour nos fins, ceci signifie que les polygones rouges représentent des zones plutôt conservatrices et, par conséquent, pourvues d'une certaine autonomie soit *scripturale* (cartes 36-37, 39, 41), soit *dialectale* (cartes 38, 40, 42), alors que les polygones bleus signalent la présence d'un brassage scriptural *en voie de gestation* au cours des 12<sup>e</sup> et 13<sup>e</sup> siècles (cartes 36-37, 39, 41) ou celle d'un brassage linguistique *accompli* à la fin du 19<sup>e</sup> siècle (cartes 38, 40, 42).

Les figurés circulaires en blanc superposés symbolisent cet état des choses: lente formation d'une 'foire scripturale' pan-oïlique au Moyen Âge, accomplissement du refoulement des derniers 'opposants' dialectaux aux périphéries septentrionale et orientale à la fin du 19<sup>e</sup> siècle. Évidemment, cette 'explosion' est due – une fois de plus – à l'expansion circulaire du type linguistique d'oïl, survenue entre les 13<sup>e</sup> et 19<sup>e</sup> siècles.

Comme les données de l'atlas de Dees de 1980 permettent une analyse-DM séparée, suivant les catégories du *vocalisme* et du *consonantisme*, et que la même procédure est possible aussi pour les données tirées de l'ALF, deux comparaisons binaires (Dees 1980 et ALF) s'imposent: voir les cartes 39-42<sup>47</sup>.

De la considération comparée des cartes 39 et 40 – relatives, toutes les deux, aux données *vocaliques* – il ressort le même effet que nous avons déjà constaté plus haut: à savoir que le caractère « pluricentrique » du domaine d'oïl médiéval est évincé, au cours des six siècles situés entre 1300 et 1900, par l'« accélération » du brassage *inter-* et *intra-*dialectal au centre et l'établissement d'une « frange de résistance » cohérente dans l'Est.

Encore un détail: sur la carte 40, le polygone rouge du point 5 constitue la dernière survivance de la large zone rouge qui, sur la carte 39, s'étend de la Bretagne romane jusqu'au Poitou et à la Saintonge.

<sup>46</sup> Ici encore, l'indice de similarité utilisé ( $SEM_{jk}$ ) est le complément de la « Distance Euclidienne Moyenne » ( $DEM_{jk}$ ):  $100 - DEM_{jk}$ .

<sup>47</sup> Voici les coordonnées de base des deux sous-corpora de l'ALF: corpus *vocalique*: 396 CT (tirées de 239 cartes-ALF originales), polynomie interne: entre 2 et 23, 2 745 taxats / aires taxatoires; corpus *consonantique*: 316 CT (tirées de 196 cartes-ALF originales), polynomie interne: entre 2 et 10, 1 068 taxats / aires taxatoires.

Il est curieux de voir que les deux cartes relatives aux données *consonantiques* (cartes 41 [scriptae du Moyen Âge] et 42 [dialectes parlés vers 1900]) offrent, en revanche, des profils choroplèthes très similaires et n'accusent pas, par conséquent, le même effet de 'restructuration' diatopique. Il en résulte que le changement linguistique postérieur au 13<sup>e</sup> siècle semble avoir connu deux vitesses: une trajectoire sensiblement plus homogène pour le *consonantisme* et une trajectoire plus mouvementée pour le *vocalisme*. Ceci signifie pour la période située entre les 13<sup>e</sup> et 19<sup>e</sup> siècles que les évolutions dans le *vocalisme*, allant dans le sens d'une plus forte homogénéisation, sont nettement plus marquées que celles dans le *consonantisme*.

#### 5.4. Comparaison dialectométrique à trois volets (Dees 1983 – Dees 1980 – ALF) D: trois analyses dendrographiques

Voir les cartes 43-45.

Parmi les méthodes exploratoires de l'EDMS, l'analyse dendrographique occupe une place de choix. Ceci tient à deux raisons:

- (1) parce que la pensée dendrographique dispose d'une longue tradition en matière de linguistique quantitative, et
- (2) parce que la grande utilité heuristique des arborescences taxométriques et de leur spatialisation est très évocatrice.

Pour une bonne compréhension des méthodes dendrographiques<sup>48</sup>, il est indiqué de considérer quatre faits très importants:

- (a) Comme la statistique moderne offre une multiplicité d'algorithmes dendrographiques, le nombre des arbres à calculer est également très grand. Néanmoins, il incombe toujours au linguiste – et jamais au statisticien – de choisir, parmi les arbres mis à sa disposition, celui qui lui paraît le plus adéquat selon ses propres présuppositions théoriques et empiriques. C'est que l'application, de la part d'un *linguiste*, de l'analyse dendrographique à un problème *linguistique*, se fait dans un contexte *linguistique* et doit donc répondre (et obéir!) uniquement à des finalités exclusivement *linguistiques*. Cela n'empêche pas qu'il est admis de choisir d'autres algorithmes dendrographiques, tout en adoptant d'autres critères de sélection, si le contexte épistémologique environnant change d'orientation.
- (b) En matière de géolinguistique, il faut toujours transposer le message classificatoire des ramifications de l'arbre ('dendrèmes') dans l'espace ('chorèmes'): d'où la nécessité absolue d'accompagner chaque arbre calculé d'une spatialisation convenable.
- (c) Du point de vue mathématique, les arbres obéissent à plusieurs impératifs qu'il ne faut pas ignorer: leur genèse statistique commence à partir des 'feuilles' (ou: objets, éléments, points d'enquête, lococlectes, etc.) et s'achève au niveau du 'tronc' (ou de la

<sup>48</sup> Pour plus de détails, cf. Goebel 2002, 29-31 et 2003, 84-89.

'racine'), alors que leur interprétation linguistique suit, en règle générale, la direction inverse.

- (d) Les ramifications (ou : bifurcations) de l'écrasante majorité des arbres créés par voie taxométrique sont exclusivement *binaires*.

L'algorithme utilisé pour la génération des arbres des cartes 43-45 a été proposé, en 1963, par le statisticien américain Joe H. Ward Jr. (1926-2011). Il appartient à la classe des méthodes dites de 'classification ascendante hiérarchique' (CAH)<sup>49</sup>. La métaphore de l'*ascension* se réfère au fait que les fusions agglomératives – qui s'opèrent, toujours par paires et selon certaines contraintes mathématiques, entre les éléments des données sujettes à classification – débutent au niveau des « feuilles » et continuent de monter vers la « racine » de l'arbre tout en générant, chemin faisant, une hiérarchie enveloppante.

L'algorithme de Ward a l'avantage de créer des classes à la fois très homogènes à l'intérieur (par la minimisation de l'*'intra-class-distance'*) et disposant de limites bien déterminées vers l'extérieur (par la maximisation de l'*'inter-class-distance'*). Leur utilité *linguistique* est très grande.

Sur les cartes 43-45, nous avons colorié, à proximité des racines des arbres en question, six groupements majeurs qui, à l'intérieur des arbres, s'appellent 'dendèmes' (D) et, après leur transfert sur la carte, 'chorèmes' (C). La finesse du coloriage, limitée à seulement six teintes, répond à des fins pratiques; à l'aide du programme VDM, elle pourrait facilement être augmentée. Les six chorèmes coloriés reflètent toutes les zones dialectales majeures du domaine d'oïl. Leur coloriage identique a deux avantages: (a) celui de mieux pouvoir identifier les différents dendèmes et chorèmes, et (b) celui de mieux pouvoir comparer les trois spatialisations entre elles.

Il ne suffit cependant pas de comparer entre elles les seules spatialisations tout en négligeant de considérer aussi la structuration interne des arbres respectifs. Que l'on jette, à ce propos, un regard sur le premier embranchement! Dans les trois cas, la bifurcation *supérieure* rassemble des ramifications (dendèmes) coloriées par des couleurs 'chaudes' (rouge, orange et jaune), alors que les dendèmes dépendant de la ramification *inférieure*, sont tous coloriés par des teintes 'froides' (vert, bleu moyen et bleu foncé).

La spatialisation des deux méga-dendèmes supérieurs génère, sur les cartes 43 et 44, une bipartition du domaine d'oïl entre l'ouest et l'est. Le même constat vaut, grosso modo, aussi pour la carte 45, avec la particularité que les chorèmes en orange et jaune des cartes 43 et 44 se sont considérablement

<sup>49</sup> Pour une description détaillée de cet algorithme, cf. Sneath / Sokal 1973, p. 283.

agrandis, sur la carte 45, vers l'est. Ceci est dû, une fois de plus, au rayonnement tous-azimuts pluriséculaire des parlers centraux du domaine d'oïl. Le même dynamisme a provoqué la réorganisation et la fusion des trois entités dialectales majeures de l'Est (picardo-wallon, lorraino-champenois et bourguignon-franc-comtois).

### 5.6. Comparaison dialectométrique à trois volets (Dees 1983 – Dees 1980 – ALF) E: cinq cartes à corrélations

Voir les cartes 46-50.

La dialectométrie corrélative (DMC) est une autre 'pièce forte' dans la panoplie méthodologique de l'EDMS. C'est qu'elle permet d'étudier la logique spatiale avec laquelle différentes sortes de gestions de l'espace – *linguistique* (*phonétique, lexicale*), *géographique*, etc. – con- ou divergent. Évidemment, le pilier théorique central de la DMC est le concept de la gestion de l'espace de la part d'un acteur, que ce soient les habitants de l'espace examiné (gestion *anthropique* ou *empirique*) ou les impératifs géométriques du théorème de Pythagore (gestion *théorique*).

Quant à la gestion *empirique* de l'espace, nous savons, depuis l'avènement de la DM<sup>50</sup> et l'apparition des premières cartes de similarité<sup>51</sup>, que les scores de similarités, relevés entre tel point de référence (j) et tous les autres points (k) du réseau examiné, diminuent toujours avec l'accroissement de la distance kilométrique<sup>52</sup> entre le point de référence et les points comparés.

Cette diminution, dont les modalités géographiques dépendent de l'emplacement du point de référence choisi, de la nature (c.-à-d. de la catégorie linguistique) des données géolinguistiques en passe d'être mesurées, et de l'indice de similarité appliqué, peut être comparée, métaphoriquement parlant, soit avec l'étalement d'une espèce d' 'énergie' dans l'espace, soit avec le versement d'un 'liquide' sur une superficie plate.

Or, il est très intéressant – du point de vue géolinguistique et aussi diachronique – d'analyser les ressemblances spatiales qui se manifestent entre ces différentes sortes d' 'étalement' ou de 'versement'. Étant donné que le langage

<sup>50</sup> Voir Séguy 1971 (*passim*).

<sup>51</sup> Voir Goebel 1976 (figures 5-7) et 1981 (figures 9-13).

<sup>52</sup> Dans ce contexte, il vaut mieux remplacer le concept de la *distance* (kilométrique) par celui de la *proximité* qui s'instaure entre le point de référence (j) et le point d'atlas comparé (k). C'est ainsi que nous sommes en présence de l'évolution parallèle de deux effets de diminution progressive dans l'espace: de la *similarité linguistique* et de la *proximité géographique* (*ProxGéo*) par rapport à un point de référence pré-établi.

de l'espace fournit toujours des informations *historiques* très utiles, voire précieuses, l'application réfléchie de la DMC constitue un outil de recherche de toute première importance.

Les cartes 46-48 se réfèrent à la corrélation spatiale qui existe, d'un côté, entre la gestion *scripturale* (cartes 46 et 47) ou *dialectale* (carte 48) du domaine d'oïl et la gestion *euclidienne* du même espace, de l'autre. La mise en relation de ces deux dimensions dépend de l'hypothèse que la propagation des similarités linguistiques dans l'espace obéit – bien que d'une manière toujours imparfaite – à certaines régularités mathématiques, et que le meilleur moyen pour découvrir ces imperfections est de les confronter avec le modèle géométrique 'parfait' du théorème de Pythagore.

Sur les cartes 46-48, la couleur *rouge* signale une grande convergence quantitative (en termes de l'indice de corrélation utilisé [ $r(\text{BP})$ ]) entre les gestions *linguistique* et *géométrique* de l'espace, alors que la couleur *bleu foncé* renvoie, au contraire, à une grande divergence (ou même dissonance) entre les deux manières de gérer l'espace.

Quelle est donc la signification de ces faits pour les objectifs de la géolinguistique ? Dans ce contexte, le constat de la dissonance – rendue visible par les plages polygonales en *bleu foncé* – est le plus suggestif. Il renvoie à des dynamismes (ou changements) linguistiques 'extra-ordinaires', c.-à-d. survenus en dehors du cadre défini par le théorème (égalisant) de Pythagore. Il est très intéressant de voir que, sur les trois cartes, les zones 'dynamiques' se situent pratiquement au même endroit, à savoir dans l'est du domaine d'oïl, qui, de ce fait, s'avère comme ayant été, dès le Moyen Âge, le moteur propulseur de l'évolution linguistique oïlique.

Il est non moins intéressant de vérifier la même convergence pour l'ouest du domaine d'oïl par le biais des polygones coloriés en rouge, orange et jaune, qui, elle, se place en revanche sous le signe d'un changement linguistique survenu plutôt 'au ralenti' et en conformité avec les impératifs géométriques de l'espace.

Ajoutons encore que la grande ressemblance entre les profils choroplèthes des cartes 46 et 47, basées, toutes les deux, sur des matériaux *scripturaux* du Moyen Âge, est moins surprenante que celle qui s'instaure entre elles et la carte 48, basée sur les données *dialectales* (élaguées) de l'ALF, collectées à la fin du 19<sup>e</sup> siècle, c.-à-d. à une distance de six siècles.

Évidemment, la leçon 'dynamique' des cartes 46-48 devrait encore être décantée dans les détails avec nos connaissances 'canoniques' de l'histoire de la langue française et des patois d'oïl.



Les cartes 49 et 50 constituent une autre application de la DMC qui s'impose d'elle-même dans le contexte d'une étude scriptologique: à savoir la comparaison corrélatrice entre les gestions *scripturale* et *dialectale* du même espace (c.-à-d. du domaine d'oïl). La carte 49 se réfère à la corrélation entre les données médiévales *littéraires* (selon Dees 1983) et celles de l'ALF, alors que la carte 50 montre la corrélation spatiale qui s'instaure entre les données médiévales *non littéraires* (selon Dees 1980) et les matériaux dialectaux tirés de l'ALF.

Une fois de plus, la grande ressemblance entre les deux profils est épataante. Les plages *bleues* symbolisent les endroits de plus grandes 'frictions' (ou: 'divergences' en profondeur) entre la gestion de l'espace pratiquée, d'une part, par les *scribes médiévaux* et, de l'autre, par les *locuteurs dialectophones* de la fin du 19<sup>e</sup> siècle. Ces frictions ou divergences s'expliquent par le brassage géolinguistique engendré au cours de l'expansion circulaire bien connue des parlers centraux du domaine d'oïl (i.e. du francien et du français « du Roy »).

Les plages en *rouge, orange et jaune*, par contre, évoquent une plus grande conformité géolinguistique entre les structures de profondeur de l'écrit médiéval et de *l'oralité* dialectale du temps de l'ALF. Le caractère 'contestataire' et 'vigoureux' des régions orientales ressort donc également de cette représentation.

## 6. Épilogue

Cet article et les recherches y ayant trait offrent plusieurs particularités:

- par la double origine géographique (c.-à-d. Amsterdam et Salzbourg) de ses fondements méthodologiques et le caractère fortuit ('trouvaille de Lauterbad') de la rencontre entre les deux;
- par le caractère relayé de la coopération entre Amsterdam (Dees [†2001], van Reenen) et Salzbourg (Goebel, Schiltz, Sobota, Smečka);
- par la double nature diamésique (textes médiévaux et dialectes modernes) des données analysées;
- par l'ampleur (1300-1900) de la fourchette diachronique desservie.

Une autre particularité réside dans la nature foncièrement *quantitative* des méthodes utilisées – incontournable d'ailleurs pour le traitement adéquat de données de masse ('big data') – et la persévérance avec laquelle tous les résultats calculés sont traduits en cartes thématiques multicolores. Ce faisant, nous avons suivi le modèle de Jules Gilliéron et d'autres pionniers de la géographie linguistique qui, moyennant la mise en carte systématique de leurs données diatopiques, ont conféré une nouvelle dimension à ces dernières tout

en ouvrant considérablement l'éventail méthodologique de la linguistique du temps.

Il s'est avéré que les méthodes de mesure conçues et développées par A. Dees pour la saisie de la polymorphie interne des textes médiévaux et de leur variabilité diatopique peuvent utilement être mariées<sup>53</sup> avec celles de la dialectométrie salzbourgeoise, et que les acquis issus de cette heureuse alliance sont d'une grande importance pour différents aspects de l'histoire de la Galloromania.

L'indéniable succès du dépouillement global des données scripturales rassemblées par Dees pour le domaine d'oïl jusqu'à la fin du 13<sup>e</sup> siècle constitue, selon nous, un encouragement pour la relance de la même expérience globalisante pour les 14<sup>e</sup> et 15<sup>e</sup> siècles, et aussi pour la couverture analogue d'autres espaces de la Romania<sup>54</sup>. Bien que l'écrit *non littéraire* et *littéraire* d'oïl, dont les dimensions prennent des allures carrément 'industrielles' aux 14<sup>e</sup> et 15<sup>e</sup> siècles, se soit déjà largement dérégionalisé et aligné sur le modèle du français central au cours de cette période, il serait néanmoins très intéressant de connaître les détails – diatopiques, diatextuels et autres – de ce gigantesque processus de remodelage, caractérisé d'un côté par une homogénéisation continue et aussi, de l'autre, par un chassé-croisé de phénomènes de déclin, de retraite et de transformation auxquels ont été sujets d'innombrables traits linguistiques en passe de devenir 'désuets'.

À la lumière de ce qui vient d'être présenté, nous sommes fermement convaincus que « le jeu en vaudrait la chandelle ».

Université de Salzbourg

Hans GOEBL  
Pavel SMEČKA

### Remerciements

- saisie électronique des résultats des calculs de corrélation effectués en 1983 par A. Dees, et confection de la figure 2: Pavel Smečka
- confection des cartes 1-50: Dumitru Kihăi, Strasbourg

<sup>53</sup> Cela n'empêche pas qu'il n'est pas du tout sûr qu'Anthonij Dees aurait apprécié les fruits de ce 'mariage', s'il avait eu la chance de les voir. Mon impression est que, dans sa pensée quantitative, l'appréciation des chiffres purs l'emportait de loin sur celle de cartes ou de structures géographiques, alors qu'au sein de l'EDMS c'est l'inverse.

<sup>54</sup> Un modèle en la matière est l'entreprise de Paul Videsott pour l'Italie septentrionale (2009).

- confection de la figure 1 et du symbolisme blanc superposé aux cartes 36-42 : Werner Goebel, Vienne
- toilettage stylistique de notre texte français : Stéphanie Lescure, Marbourg sur Lahn (Allemagne)

L'expression de ma parfaite reconnaissance va à toutes les personnes mentionnées pour leur précieuse et patiente collaboration (HG).

## 7. Abréviations techniques souvent utilisées

AIS	Atlante italo-svizzero [Jaberg / Jud 1926-1940]	<i>passim</i>
ALF	Atlas linguistique de la France	<i>passim</i>
AT	aire taxatoire	cf. le paragraphe 5.3.2.
C	chorème	cf. le paragraphe 5.4.
CAF	coefficient d'asymétrie de [Ronald Aylmer] Fisher (1890-1962)	cf. le paragraphe 5.3.2.
CAH	Classification ascendante hiérarchique	cf. le paragraphe 5.4.
D	dendrème	cf. le paragraphe 5.4.
DC	dendrème et chorème	cf. le paragraphe 5.4.
Dees 1983	ensemble des calculs de localisation effectués par Anthonij Dees en 1983 à Amsterdam («trouvaille de Lauterbad»)	cf. le paragraphe 2.
DEM <sub>jk</sub>	Distance Euclidienne Moyenne (entre les points d'atlas <i>j</i> et <i>k</i> )	voir les cartes 30-31, 33-34, 36-37 et 39 et aussi les paragraphes 5.3. et 5.4.
DM	dialectométrie, relatif à la dialectométrie, dialectométrique	<i>passim</i>
DMC	dialectométrie corrélative	cf. le paragraphe 5.6.
EDMS	École Dialectométrique de Salzbourg	<i>passim</i>
IPI(1) <sub>jk</sub>	Indice Pondéré d'Identité (entre les points d'atlas <i>j</i> et <i>k</i> ) (muni du poids/de la pondération 1)	voir les cartes 11, 14, 17, 18, 19, 26, 29,32, 35 38, 40,42, 45 et 48-50 et aussi le paragraphe 5.
IRI <sub>jk</sub>	Indice Relatif d'Identité (entre les points d'atlas <i>j</i> et <i>k</i> )	<i>passim</i>
MEDMW	algorithme d'intervallisation (ou de visualisation)	voir les cartes 30-32 ainsi que 36-42.

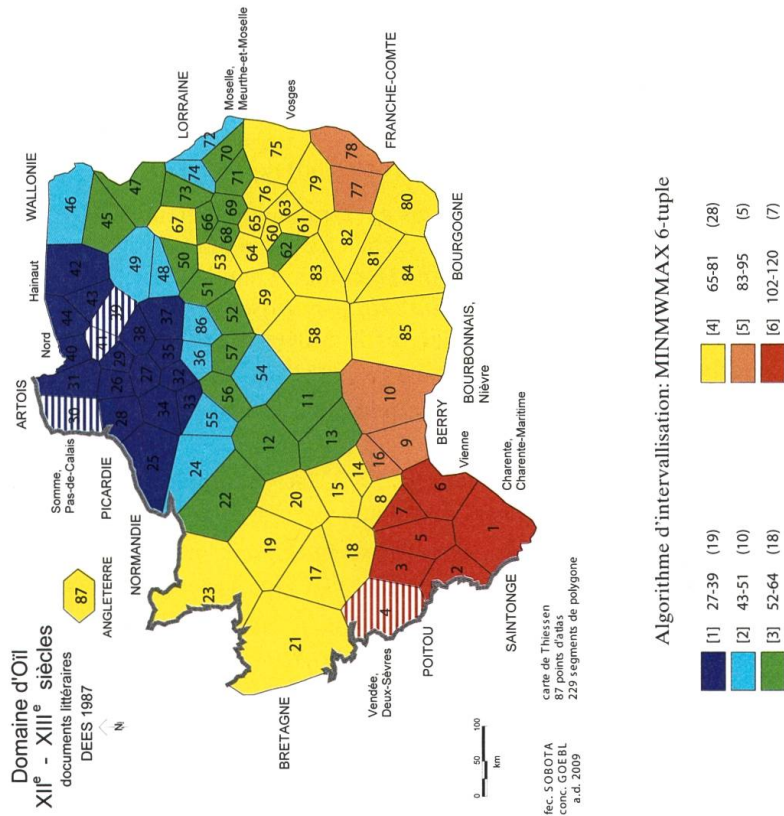
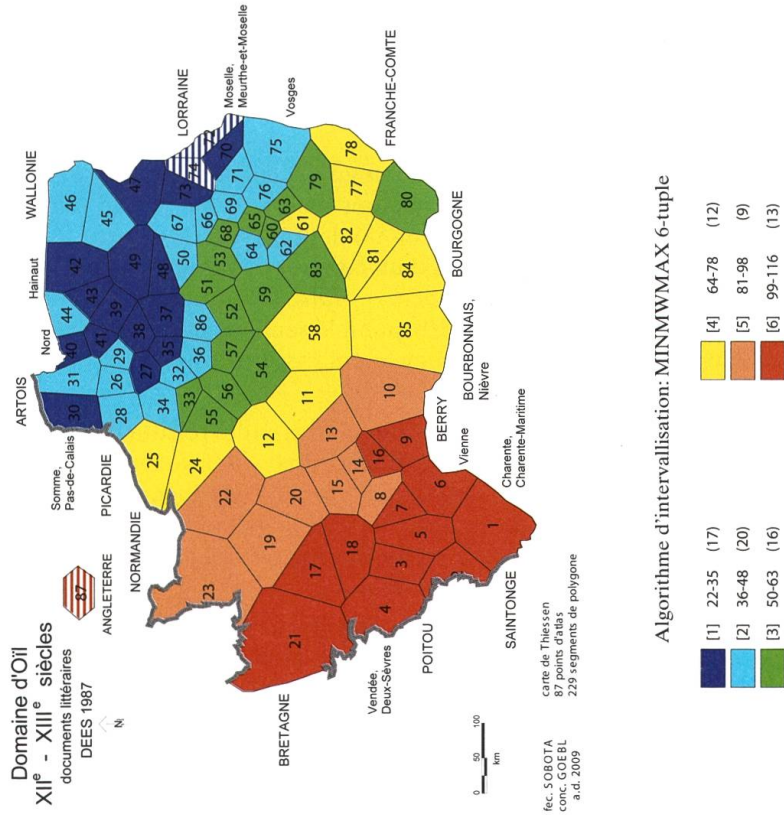
MINMWMAX	algorithme d'intervallisation (ou de visualisation)	<i>passim</i>
NCA	Nouveau Corpus d'Amsterdam	cf. le paragraphe 2.
PMK	Produkt-Moment-Korrelationskoeffizient	= r(BP)
$r(\text{BP})_{jk}$	coefficient de corrélation (entre les points d'atlas $j$ et $k$ ), proposé par [Auguste] Bravais (1811-1863) et [Karl] Pearson (1857-1936)	<i>passim</i> (= PMK)
VDM	Visual DialectoMetry (logiciel dialectométrique créé en 1999-2000 par Edgar Haimerl et utilisé couramment à Salzbourg)	<i>passim</i>

## 8. Références bibliographiques

- AIS = Jaberg, Karl / Jud, Jakob, 1928-1940. *Sprach- und Sachatlas Italiens und der Südschweiz*, 8 vols., Zofingen, Ringier (réimpression : Nendeln, Kraus 1971).
- ALF = Gilliéron, Jules / Edmont, Edmond, 1902-1910. *Atlas linguistique de la France*, 10 vols., Paris, Champion (réimpression : Bologne, 1968).
- Carles, Hélène, 2016. «La régionalité lexicale dans la Galloromania avant 1100», in: Martin Glessgen / David Trotter (éds.), *La régionalité lexicale au Moyen Âge* (Actes du colloque de Zurich, 7/8 sept. 2015), Strasbourg, ÉLiPhi, 99-110.
- Carles, Hélène, 2017. *Trésor galloroman des origines (TGO). Les trajectoires étymologiques et géolinguistiques du lexique galloroman en contexte latin (ca 800 – 1120)*, Strasbourg, ÉLiPhi (sous presse).
- Dees, Anthonij, 1980. *Atlas des formes et des constructions des chartes françaises du 13<sup>e</sup> siècle*, Tübingen, Niemeyer (*Beihefte zur ZrP*, 178).
- Dees, Anthonij, 1983. Total des calculs de localisation effectués par Anthonij Dees en 1983 à Amsterdam ('trouvaille de Lauterbad'): voir le paragraphe 2.
- Dees, Anthonij, 1984. «Regards quantitatifs sur les variations régionales en ancien français», in: Goebel, Hans (ed.), *Dialectology*, Bochum, Brockmeyer, 102-120 (*Quantitative Linguistics*, vol. 11).
- Dees, Anthonij, 1985. «Dialectes et scriptae à l'époque de l'ancien français», *RLiR* 49, 87-117.
- Dees, Anthonij, 1986. «Vers un atlas linguistique de l'ancien français écrit», in: *Actes du XVII<sup>e</sup> Congrès International de Linguistique et Philologie Romanes* (Aix-en-Provence 1983), Aix-en-Provence, Université de Provence; Marseille, Jeanne Laffitte, vol. 6, 505-517.
- Dees, Anthonij, 1987. *Atlas des formes linguistiques des textes littéraires de l'ancien français*, Tübingen, Niemeyer (*Beihefte zur ZrP*, 212).

- de Jong, Thera, 1988. «L'Anglo-normand du 13<sup>e</sup> siècle», in: van Reenen, Piet / van Reenen-Stein, Karin (ed.), *Distributions spatiales et temporelles. Constellations des manuscrits. Études de variation linguistique offertes à Anthonij Dees à l'occasion de son 60<sup>e</sup> anniversaire. Spatial and temporal distributions. Manuscript constellations. Studies in language variation offered to Anthonij Dees on the occasion of his 60th birthday*, Amsterdam, Philadelphia, Benjamins, 103-112.
- de Jong, Thera, 1996. «Anglo-French in the 13th and 14th Centuries: Continental or Insular Dialect?», in: Nielsen, Hans F. / Schøsler, Lene (ed.), *The Origins and Development of Emigrant Languages*. Proceedings from the Second Rasmus Rask Colloquium (Odense University, November 1994), Odense, University Press, 55-70.
- Goebel, Hans, 1976. «La dialectométrie appliquée à l'ALF (Normandie)» in: *Atti del XIV Congresso internazionale di linguistica e filologia romanza* (Napoli, 15-20 aprile 1974), Naples, Macchiaroli, Amsterdam, Benjamins, vol. 2, 165-195.
- Goebel, Hans, 1984. *Dialektometrische Studien. Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, Tübingen, Niemeyer, 3 vols. (*Beihefte zur ZrP*, 191-193).
- Goebel, Hans, 1987. «Points chauds de l'analyse dialectométrique: pondération et visualisation», *RLiR* 51, 63-118.
- Goebel, Hans, 1998. «Zu einer dialektometrischen Analyse der Daten des Dees-Atlases von 1980», in: Werner, Edeltraud / Liver, Ricarda / Stork, Yvonne / Nicklaus, Martina (ed.), *Et multum et multa. Festschrift für Peter Wunderli zum 60. Geburtstag*, Tübingen, Narr, 293-309.
- Goebel, Hans, 2001. «La strutturazione geolinguistica del dominio d'Oïl nel 13<sup>o</sup> secolo e alla fine dell'Ottocento: un raffronto dialettometrico», in: Zamboni, Alberto / Del Puente, Patrizia / Vigolo, Maria Teresa (ed.), *La dialettologia oggi fra tradizione e nuove metodologie*. Atti del Convegno Internazionale (Pisa 2000), Alessandria, Edizioni ETS, 11-43.
- Goebel, Hans, 2002. «Analyse dialectométrique des structures de profondeur de l'ALF», *RLiR* 66, 5-63 (avec 24 cartes en couleurs).
- Goebel, Hans, 2003. «Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur», *Estudis Romànics* 25, 59-96 (avec 24 cartes en couleurs).
- Goebel, Hans, 2005. «La dialectométrie corrélative. Un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme», *RLiR* 69, 321-367 (avec 24 cartes en couleurs).
- Goebel, Hans, 2006. «Sur le changement macrolinguistique survenu entre 1300 et 1900 dans le domaine d'oïl. Une étude diachronique d'inspiration dialectométrique», *Linguistica* 46, 3-43 (avec 12 cartes en couleurs).
- Goebel, Hans, 2008a. «Sur le changement macrolinguistique survenu entre 1300 et 1900 dans le domaine d'oïl. Une étude diachronique d'inspiration dialectométrique», *Dialectologia* (Barcelona) 1, 3-43 [version remaniée et corrigée de Goebel 2006] (avec 24 cartes en couleurs).
- Goebel, Hans, 2008b. «La dialettometrizzazione integrale dell'AIS. Presentazione dei primi risultati», *RLiR* 72, 25-113 (avec 48 cartes en couleurs).

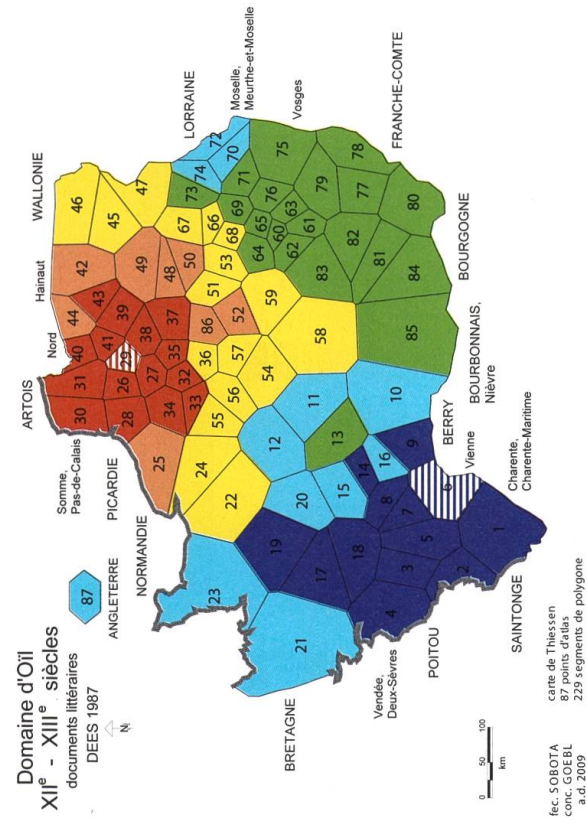
- Goebel, Hans, 2011. «Die beiden Skripta-Atlanten von Anthonij Dees (1980 und 1987) im dialektometrischen Vergleich: eine erste Bilanz», in: Overbeck, Anja / Schweickard, Wolfgang / Völker, Harald (ed.), *Lexikon, Varietät, Philologie. Romanistische Studien. Günter Holtus zum 65. Geburtstag*, Berlin, Boston, Walter de Gruyter, 665-677 (avec 8 cartes en couleurs).
- Goebel, Hans, 2012. «L'aménagement scripturaire du domaine d'oïl médiéval à la lumière des calculs de localisation d'Anthonij Dees effectués en 1983: une étude d'inspiration scriptométrique», 11 p. (avec 24 cartes en couleurs). (Version écrite d'une conférence tenue à Venise en 2011 à l'occasion d'un congrès de la revue «Medioevo romanzo», dédié au sujet suivant: *Il problema della scripta. Gli antichi testi romanzi tra filologia, dialettologia e storia della lingua*; disponible seulement en ligne: <[www.medioevoromanzo.it/uploads/H.Goebel\\_Amenagement\\_scripturaire\\_2012\\_article.pdf](http://www.medioevoromanzo.it/uploads/H.Goebel_Amenagement_scripturaire_2012_article.pdf)>)
- Goebel, Hans, 2014. «L'impact de la polynymie des cartes d'atlas sur le résultat de calculs dialectométriques», in: Polska Akademia Umiejetnosci. Instytut Filologii Romanskiej Uniwersytetu Jagiellonskiego (ed.), *Linguistique romane et Linguistique indoeuropéenne. Mélanges offerts à Witold Mańczak à l'occasion de son 90<sup>e</sup> anniversaire*, Cracovie, Polska Akademia Umiejetnosci. Instytut Filologii Romanskiej Uniwersytetu Jagiellonskiego, 243-260 (avec 10 cartes en couleurs).
- Goebel, Hans / Schiltz, Guillaume, 2001. «Der „Atlas des formes et des constructions des chartes françaises du 13<sup>e</sup> siècle“ von Anthonij Dees (1980) – dialektometrisch betrachtet», in: Gärtner, Kurt / Holtus, Günter / Rapp, Andrea / Völker, Harald (ed.), *Skripta, Schreiblandschaften und Standardisierungstendenzen. Urkundensprachen im Grenzbereich von Germania und Romania im 13. und 14. Jahrhundert*. Beiträge zum Kolloquium vom 16. bis 18. September 1998 in Trier, Trier, Kliomedia, 169-221 (avec 22 cartes en couleurs).
- Kunstmann, Pierre / Stein, Achim (ed.), 2007. *Le Nouveau Corpus d'Amsterdam. Actes de l'atelier de Lauterbad, 23-26 février 2006*, Stuttgart, Steiner (ZFSL, Beihefte, 24).
- Remacle, Louis, 1992. *La différenciation dialectale en Belgique romane avant 1600*, Genève, Droz.
- Schøsler, Lene (2002). «Nécrologie: Anthonij Dees (1928-2001)», *RLiR* 66, 614-618.
- Schøsler, Lene / van Reenen, Pieter (2000). «Corpus et stemma en ancien et en moyen français. Bilan, résultats et perspectives des recherches à l'Université libre d'Amsterdam et dans les institutions collaboratrices», in: Buridant, Claude (ed.), *Le moyen français. Le traitement du texte (édition, appareil critique, glossaire, traitement électronique)*. Actes du IX<sup>e</sup> colloque international sur le moyen français (Strasbourg, 29-31 mai 1997), Strasbourg, PUS, 25-54.
- Séguy, Jean, 1971. «La relation entre la distance spatiale et la distance lexicale», *RLiR* 35, 335-357.
- Sneath, Peter Henry Andrews / Sokal, Robert Reuven, 1973. *The Principles and Practice of Numerical Classification*, San Francisco, Freeman.
- Videsott, Paul (2009). *Padania scrittologica: analisi scrittologiche e scrittometriche di testi in italiano settentrionale antico dalle origini al 1525*, Tübingen, Niemeyer (Beihefte zur ZrP, 343).
- Wüest, Jakob, 1979. *La dialectalisation de la Gallo-Romania: problèmes phonologiques*, Berne, Francke.



Carte 2: Localisation quantitative du « Bestiaire » de Guillaume le Clerc  
Références: Dees 1983 (VDM): 182; Dees 1987, 533; Nouveau Corpus d'Amsterdam: 36

Carte 1: Localisation quantitative de « Der sogenannte poitevinische Pseudo-Turpin »  
Références: Dees 1983 (VDM): 4; Dees 1987: 519; Nouveau Corpus d'Amsterdam: 216

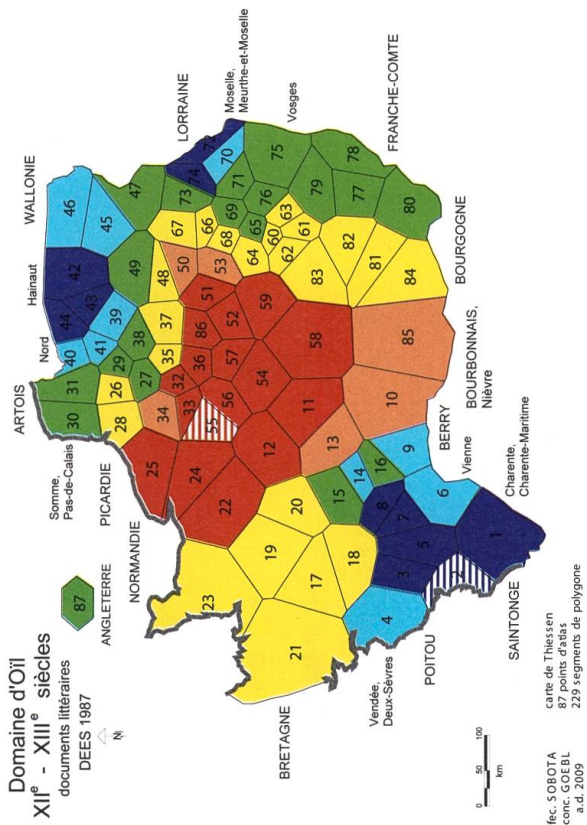
Hachures blanches: scores minimal et maximal



Algorithme d'intervallisation: MINMWMAX 6-tuple



Carte 4: Localisation quantitative du « Renart le Nouvel » par Jacquemart Gliellee  
Références: Dees 1983 (VDM): 55; Dees 1987, 523; Nouveau Corpus d'Amsterdam: 191



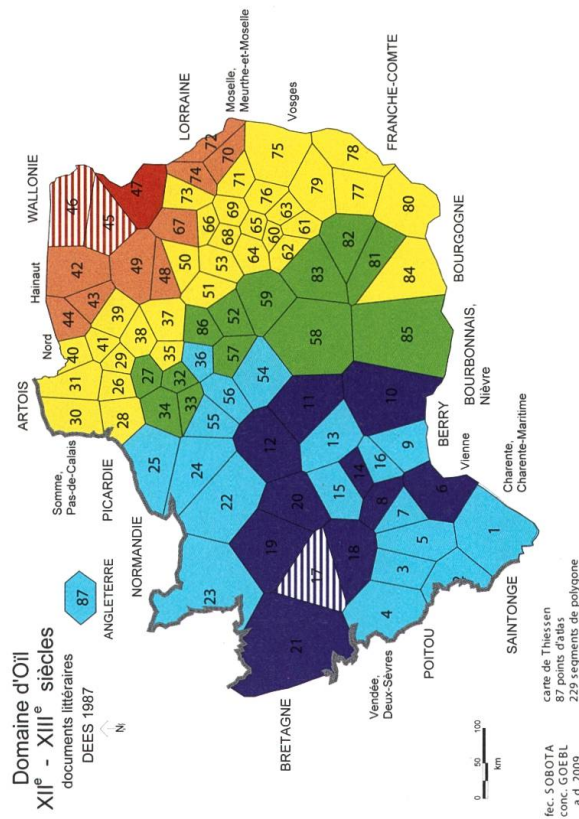
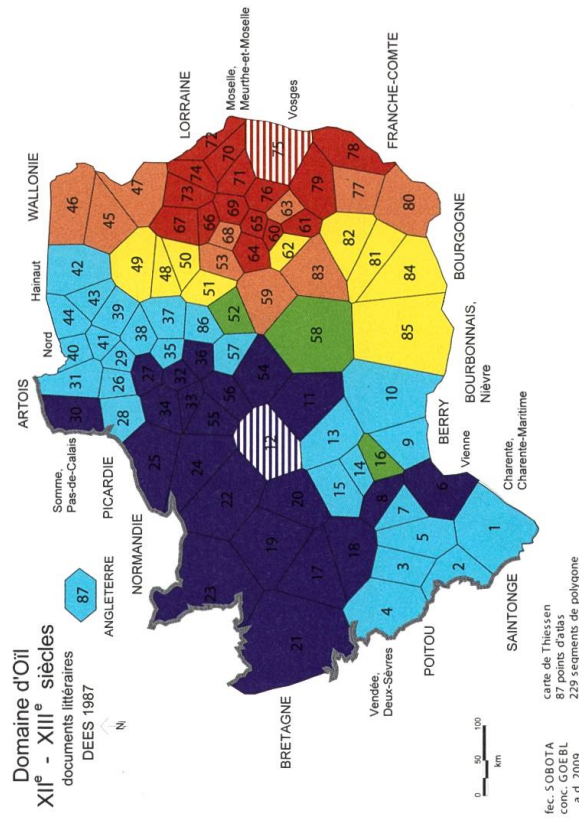
Algorithme d'intervallisation: MINMWMAX 6-tuple



Carte 3: Localisation quantitative de 2000 vers tirés du ms. U du « Perceval » de Chrétien de Troyes  
Références: Dees 1983 (VDM): 110; Dees 1987, 527; Nouveau Corpus d'Amsterdam: 214

Hachures blanches: scores minimal et maximal

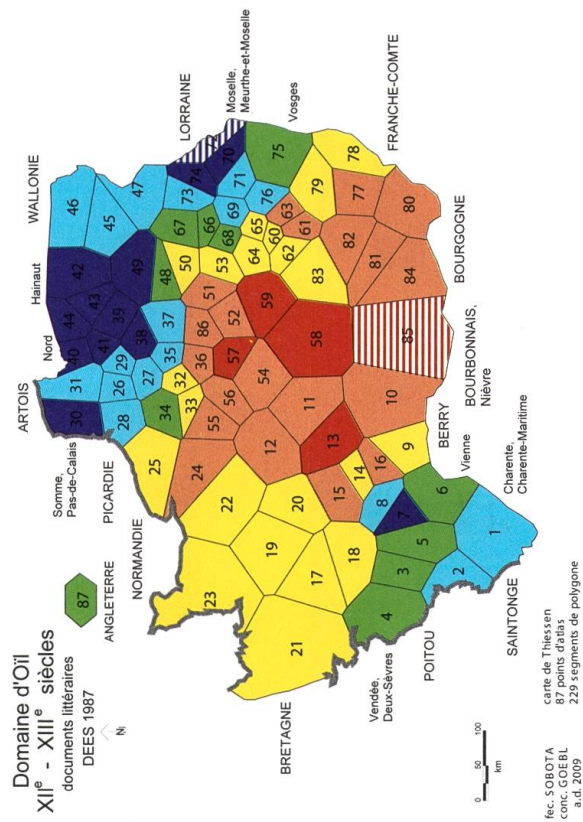




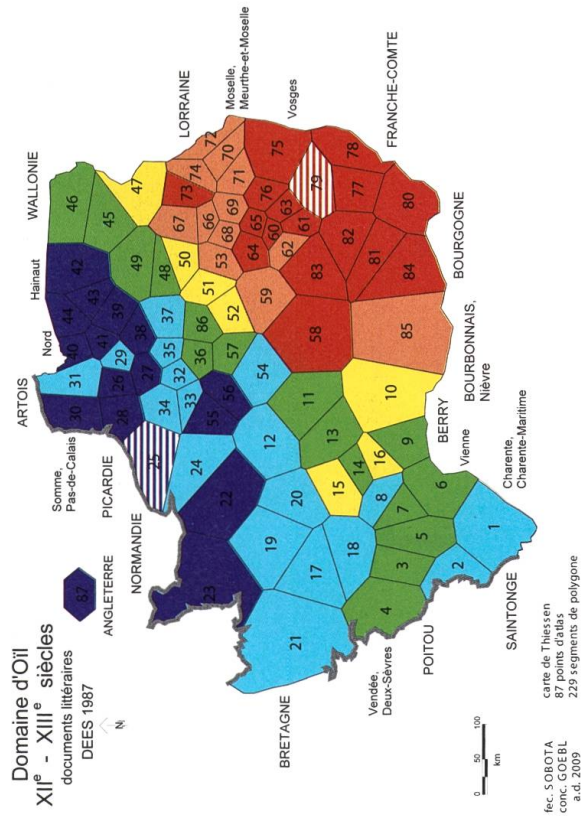
Carte 5: Localisation quantitative du « Médecinaire liégeois du XIII<sup>e</sup> siècle »  
Références: Dees 1983 (VDM): 85; Dees 1987, 525; Nouveau Corpus d'Amsterdam: 65

Hachures blanches: scores minimal et maximal

Carte 6: Localisation quantitative du « Dialogus anime conquerentis et rationis consolantis »  
Références: Dees 1983 (VDM): 153; Dees 1987: 531; Nouveau Corpus d'Amsterdam: 30



Algorithme d'intervallisation: MINNWMAX 6-tuple



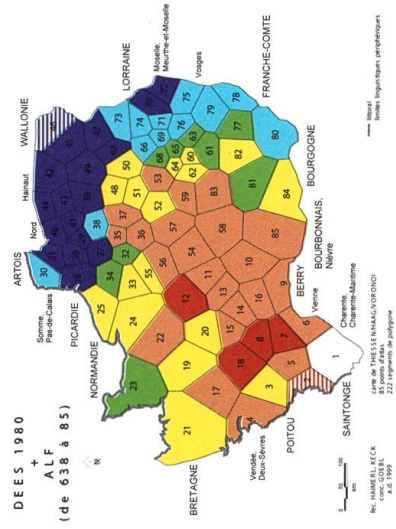
Algorithme d'intervallisation: MINNWMAX 6-tuple



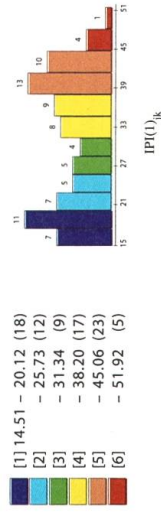
Carte 7: Localisation quantitative des « Poignes d'enfer, les deux chevaliers »  
Références: Dees 1983 (VDM): 156; Dees 1987, 531; Nouveau Corpus d'Amsterdam: 40

Hachures blanches : scores minimal et maximal

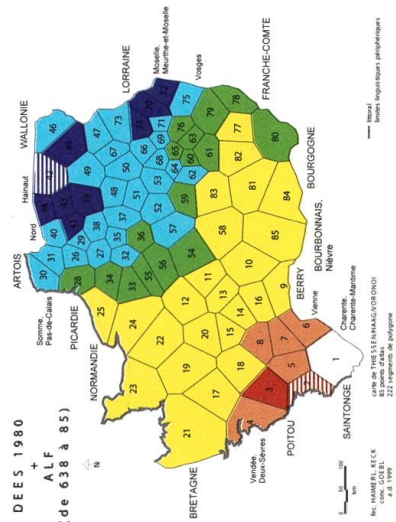
Carte 8: Localisation quantitative du « Roman de la Rose » de Guillaume de Lorris  
Références: Dees 1983 (VDM): 171; Dees 1987, 532; Nouveau Corpus d'Amsterdam: 245



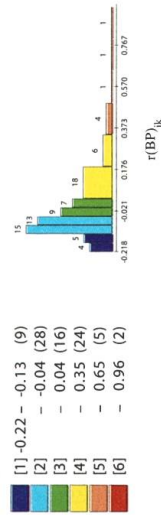
Algorithme d'intervallisation MINMWMAX 6-tuple



Carte 11: Corpus: 1 279 cartes de travail (toutes les catégories linguistiques) tirées de l'ALF  
Indice de similarité: Indice Pondéré d'Identité (avec le poids 1)  $[IP(1)_{jk}]$

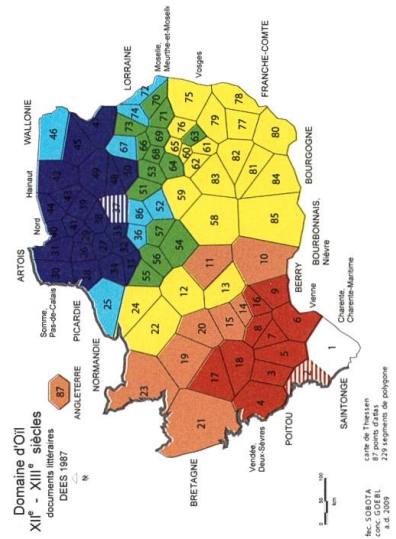


Algorithme d'intervallisation MINMWMAX 6-tuple

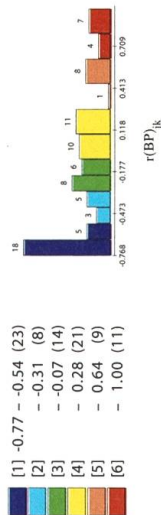


Carte 10: Corpus: 268 attributs scripturaux (Dees 1980)  
Indice de similarité: coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]$

Hachures blanches: scores minimal et maximal

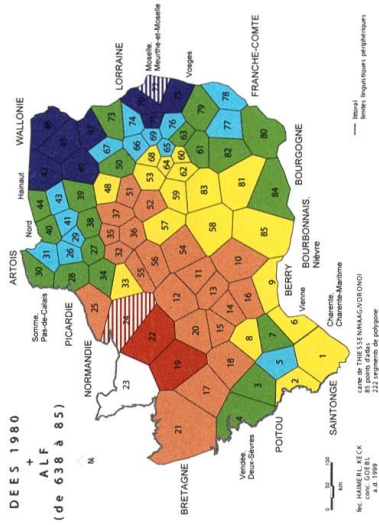


Algorithme d'intervallisation MINMWMAX 6-tuple

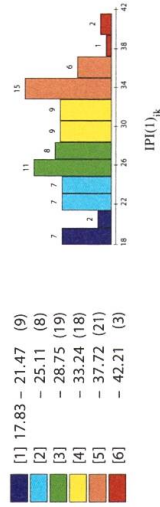


Carte 9: Corpus: 222 vecteurs de localisation (Dees 1983)  
Indice de similarité: coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]$

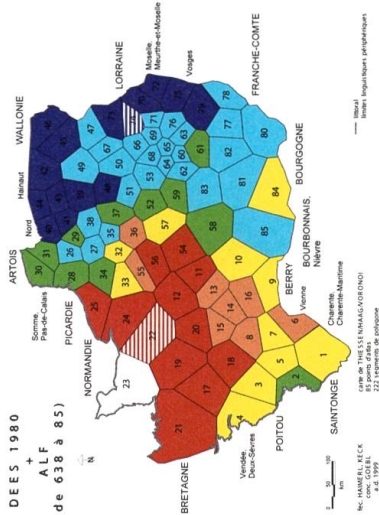
Cartes 9-11: Cartes de similarité relatives au point de référence 1 (Charente, Charente-Maritime et ALF 518, Chassors)



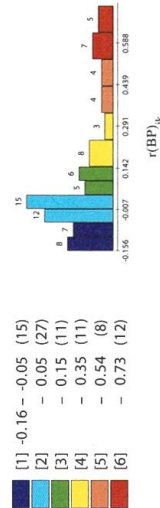
Algorithme d'intervallisation  
MINMWMAX 6-tuple



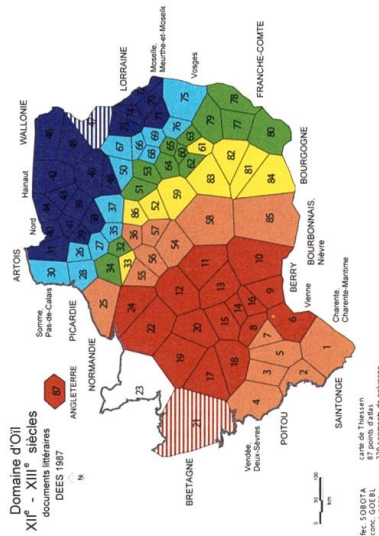
Carte 14: Corpus: 1 279 cartes de travail (toutes les catégories linguistiques) tirées de l'ALF  
Indice de similarité: Indice Pondéré d'Identité (avec le poids 1)  $[IPI(1)_{jk}]$



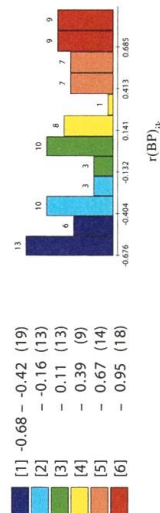
Algorithme d'intervallisation  
MINMWMAX 6-tuple



Carte 13: Corpus: 268 attributs scripturaux (Dees 1980)  
Indice de similarité: coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]$



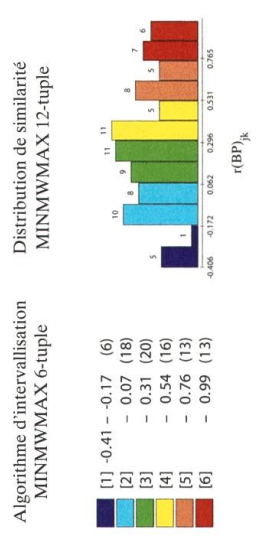
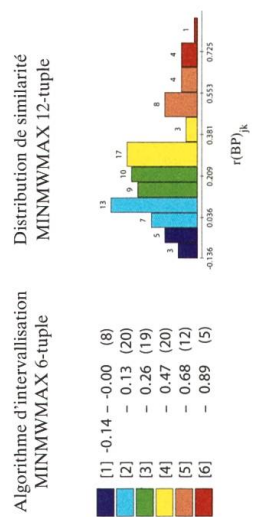
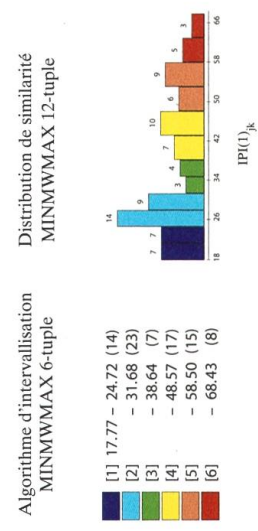
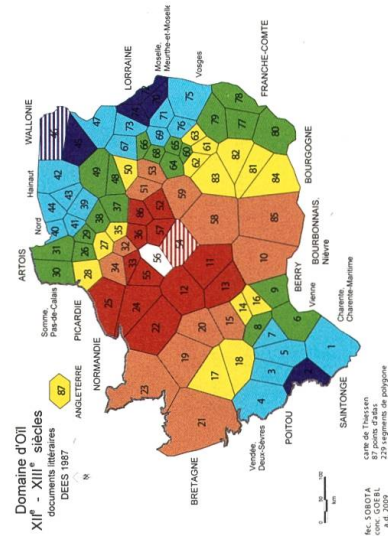
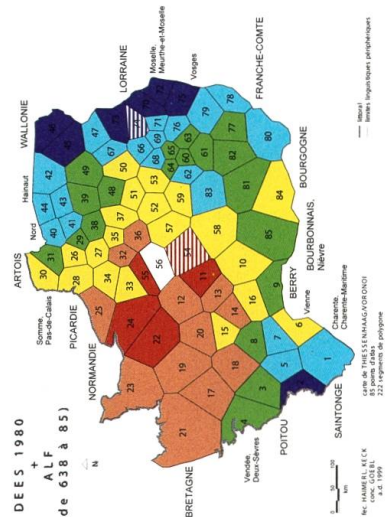
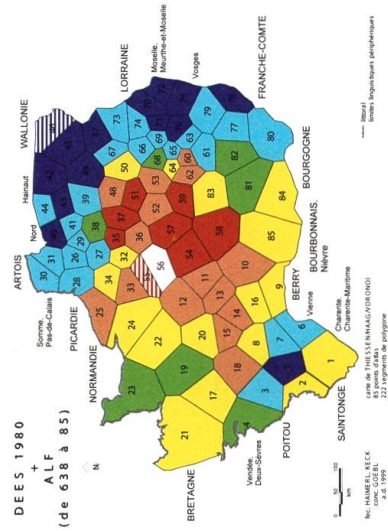
Algorithme d'intervallisation  
MINMWMAX 6-tuple



Carte 12: Corpus: 222 vecteurs de localisation (Dees 1983)  
Indice de similarité: coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]$

Hachures blanches : scores minimal et maximal

Cartes 12-14: Cartes de similarité relatives au point de référence 23 (Manche et ALF 377, Pont-Hébert)



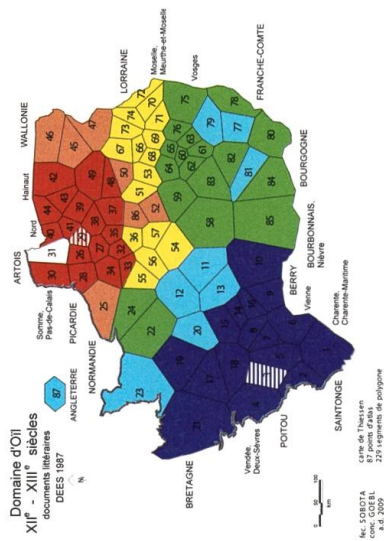
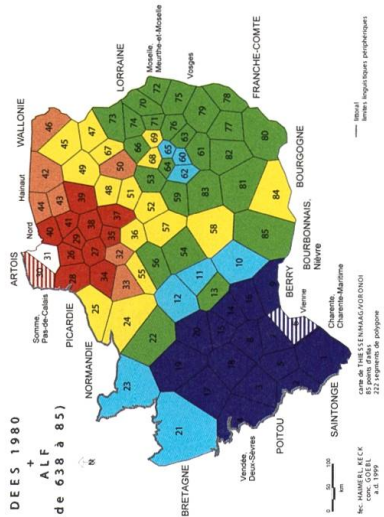
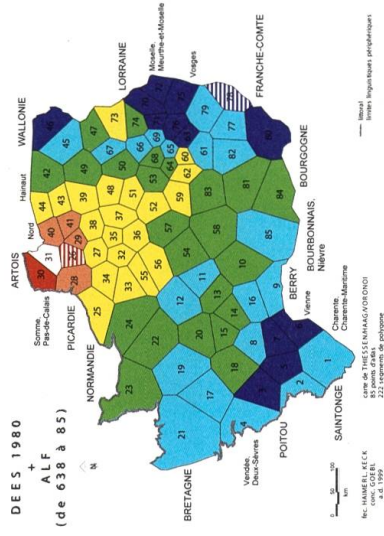
Carte 17: Corpus: 1 279 cartes de travail (toutes les catégories linguistiques) tirées de l'ALF  
Indice de similarité: Indice Pondéré d'Identité (avec le poids 1)  $[IPI(1)_{jk}]$

Carte 16: Corpus: 268 attributs scripturaux (Dees 1980)  
Indice de similarité: coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]$

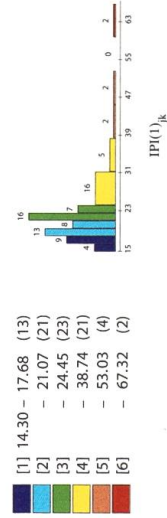
Carte 15: Corpus: 222 vecteurs de localisation (Dees 1983)  
Indice de similarité: coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]$

Hachures blanches: scores minimal et maximal

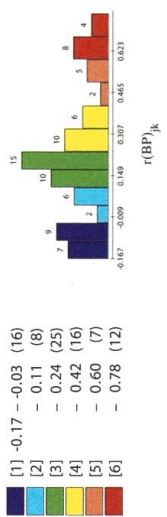
Cartes 15-17: Cartes de similarité relatives au point de référence 56 (Paris et ALF 226, Plessis-Piquet)



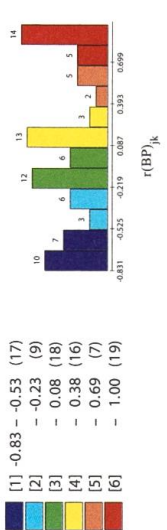
Algorithme d'intervallisation MINMWMAX 6-tuple



Algorithme d'intervallisation MINMWMAX 6-tuple



Algorithme d'intervallisation MINMWMAX 6-tuple



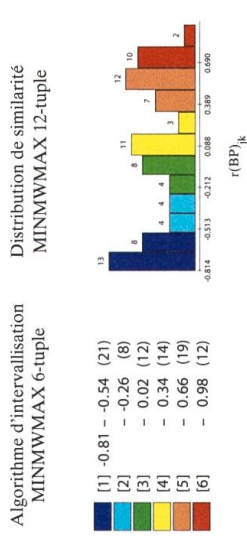
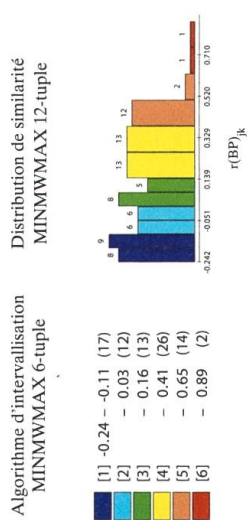
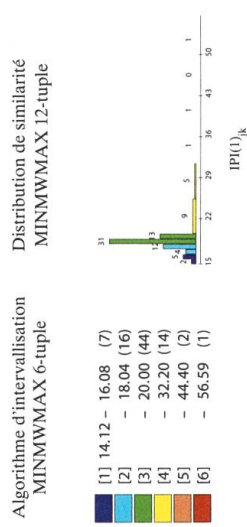
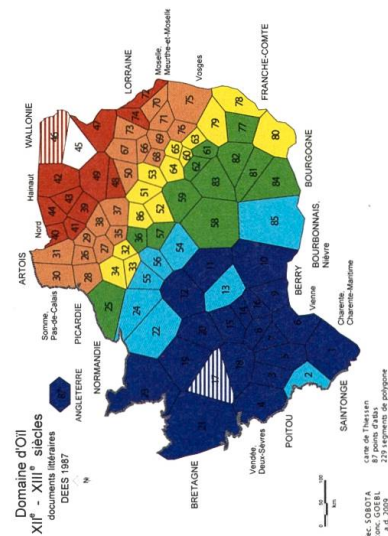
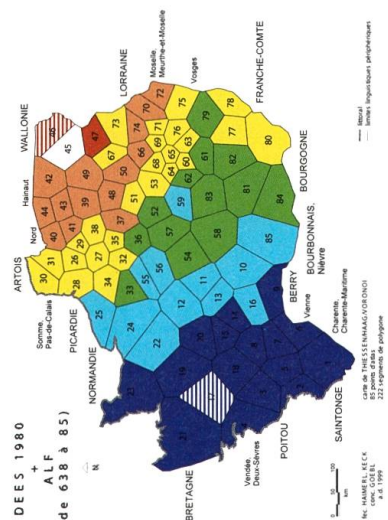
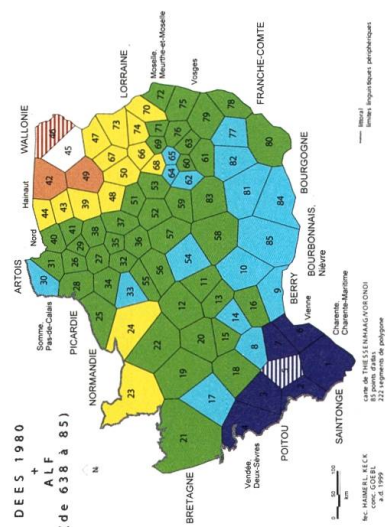
Carte 20: Corpus: 1 279 cartes de travail (toutes les catégories linguistiques) tirées de l'ALF  
Indice de similarité: Indice Pondéré d'Identité (avec le poids 1) IPI(1)<sub>jk</sub>

Carte 19: Corpus: 268 attributs scripturaux (Dees 1980)  
Indice de similarité: coefficient de corrélation de Bravais et Pearson [r(BP)<sub>jk</sub>]

Carte 18: Corpus: 222 vecteurs de localisation (Dees 1983)  
Indice de similarité: coefficient de corrélation de Bravais et Pearson [r(BP)<sub>jk</sub>]

Hachures blanches : scores minimal et maximal

Cartes 18-20 : Cartes de similarité relatives au point de référence 31 (Pas-de-Calais, Centre-Nord et ALF 296, Isbergues)



Carte 23: Corpus: 1 279 cartes de travail (toutes les catégories linguistiques) tirées de l'ALF  
Indice de similarité: Indice Pondéré d'Identité (avec le poids 1) [IPI]<sub>jk</sub>

Carte 22: Corpus: 268 attributs scripturaux (Dees 1980)  
Indice de similarité: coefficient de corrélation de Bravais et Pearson [r(BP)]<sub>jk</sub>

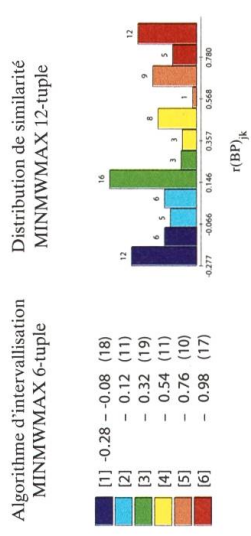
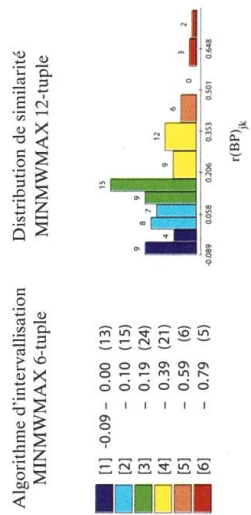
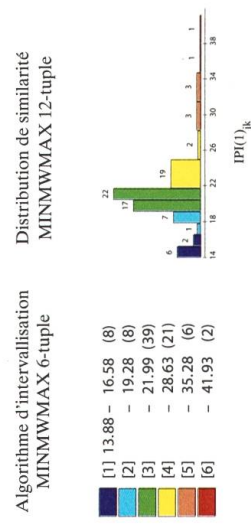
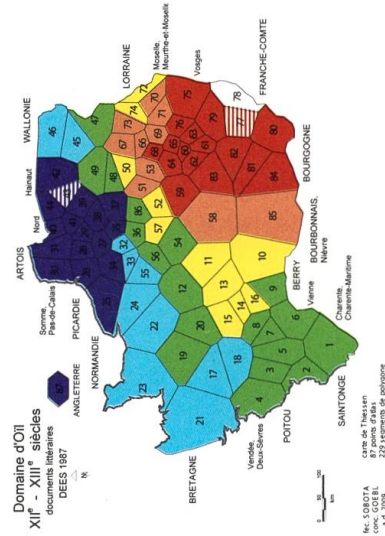
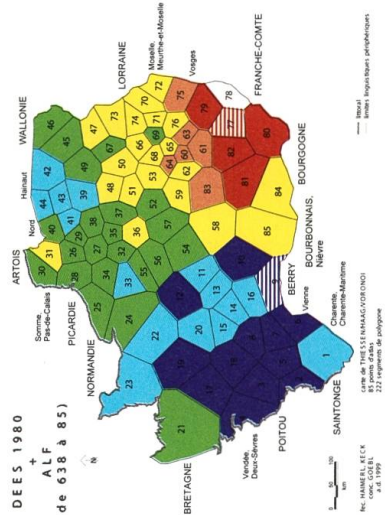
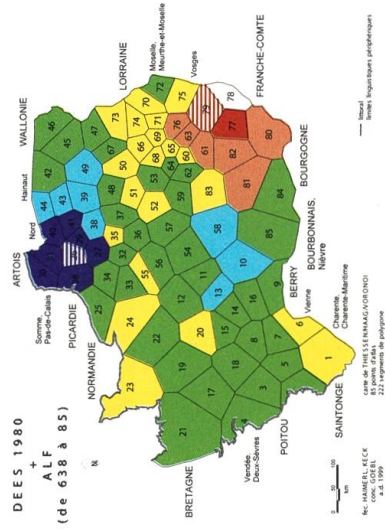
Carte 21: Corpus: 222 vecteurs de localisation (Dees 1983)  
Indice de similarité: coefficient de corrélation de Bravais et Pearson [r(BP)]<sub>jk</sub>

Hachures blanches: scores minimal et maximal

Cartes 21-23: Cartes de similarité relatives au point de référence 45 (Wallonie et ALF 192, Bomal-lez-Durbuy)







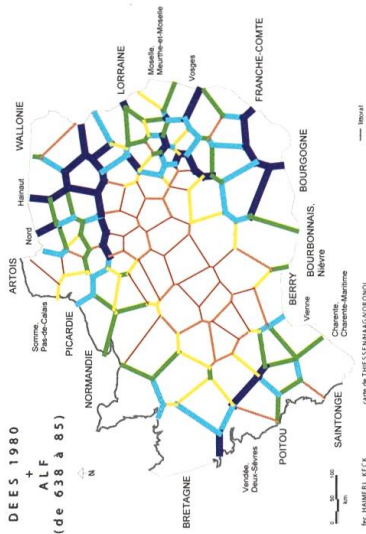
Carte 29: Corpus: 1 279 cartes de travail (toutes les catégories linguistiques) tirées de l'ALF  
Indice de similarité: Indice Pondéré d'Identité (avec le poids 1)  $[IP(1)_{jk}]$

Carte 28: Corpus: 268 attributs scripturaux (Dees 1980)  
Indice de similarité: coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]$

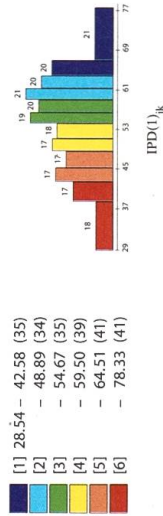
Carte 27: Corpus: 222 vecteurs de localisation (Dees 1983)  
Indice de similarité: coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]$

Hachures blanches: scores minimal et maximal

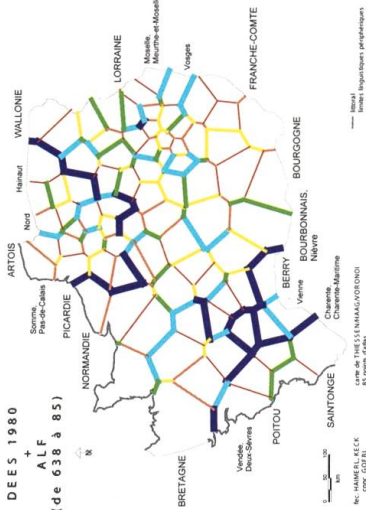
Cartes 27-29: Cartes de similarité relatives au point de référence 78 (Doubs et ALF 53, Saint-Hippolyte)



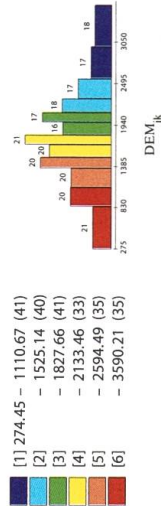
Algorithme d'intervallisation  
MEDMW 6-tuple



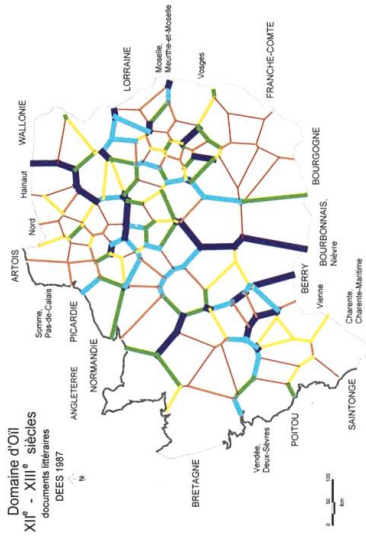
Carte 32: Corpus: 1 279 cartes de travail (toutes les catégories linguistiques) tirées de l'ALF  
Indice de distance: Indice Pondéré de Distance (avec le poids 1) (IPD(1)<sub>jk</sub>)



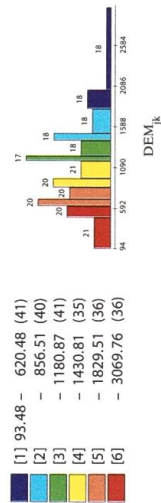
Algorithme d'intervallisation  
MEDMW 6-tuple



Carte 31: Corpus: 268 attributs scripturaux (Dees 1980)  
Indice de distance: Distance Euclidienne Moyenne (DEM<sub>jk</sub>)

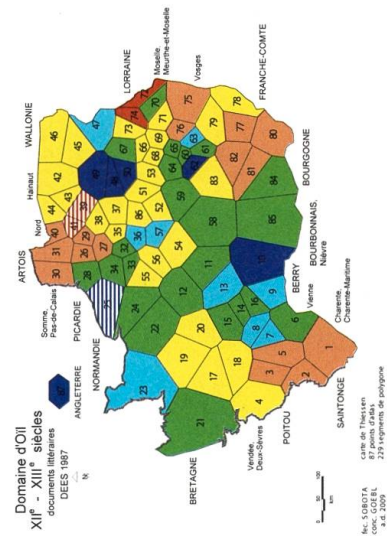
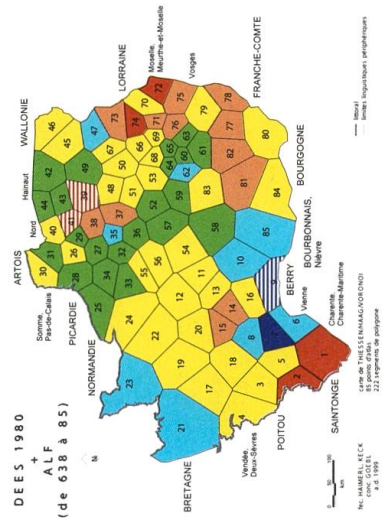
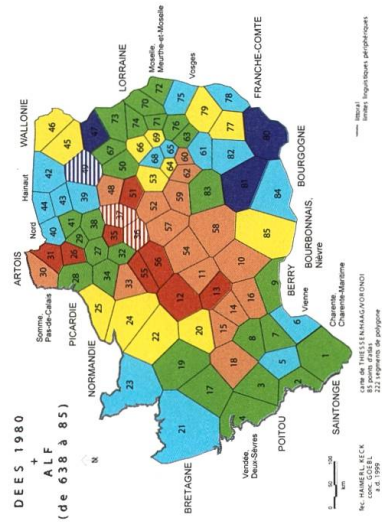


Algorithme d'intervallisation  
MEDMW 6-tuple

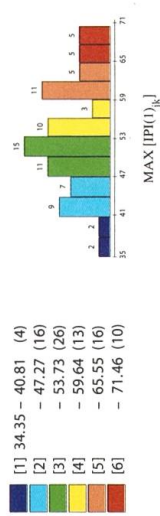


Carte 30: Corpus: 222 vecteurs de localisation (Dees 1983)  
Indice de distance: Distance Euclidienne Moyenne (DEM<sub>jk</sub>)

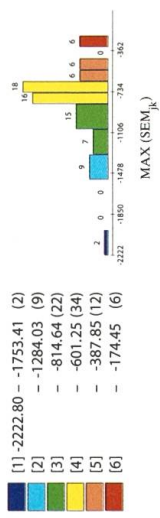
Cartes 30-32: Synthèses isoglossiques (ou: cartes à cloisons)



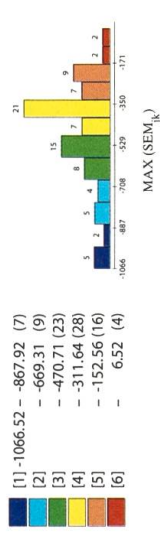
Algorithme d'intervallisation  
MINMWMAX 6-tuple



Algorithme d'intervallisation  
MINMWMAX 6-tuple



Algorithme d'intervallisation  
MINMWMAX 6-tuple



Carte 35: Corpus: 1 279 cartes de travail (toutes les catégories linguistiques) tirées de l'ALF  
Indice de similarité: Indice Pondéré d'Identité (avec le poids 1) [IP(1)\_{jk}]

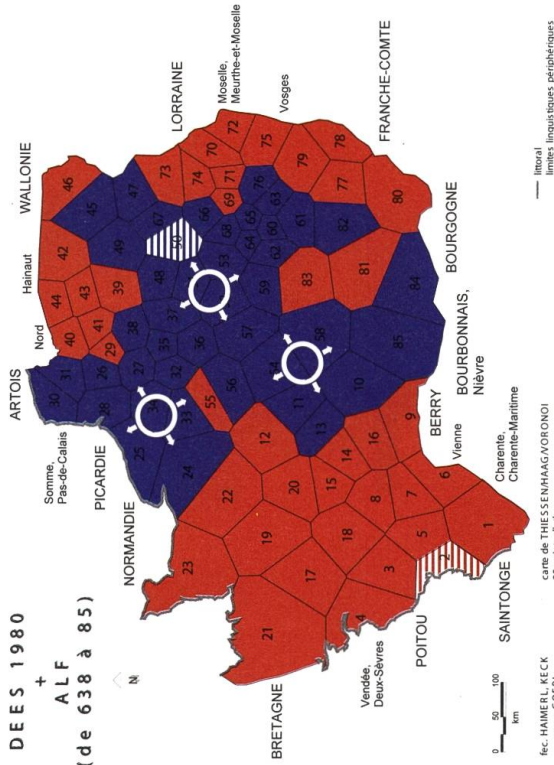
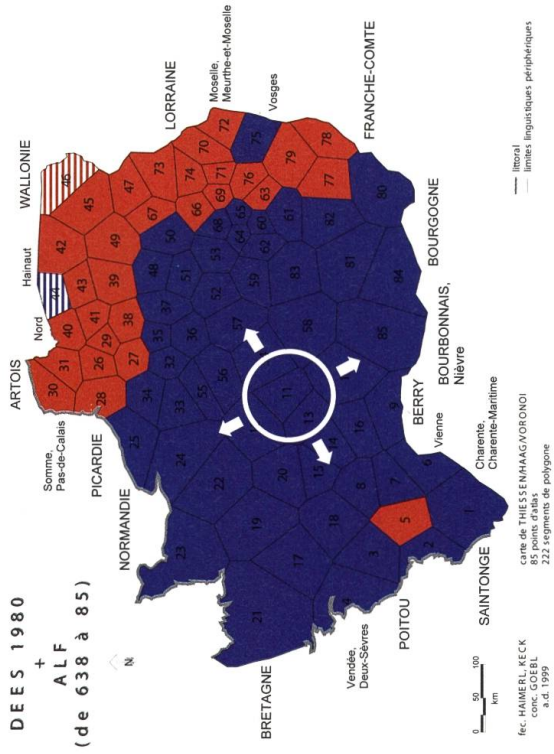
Carte 34: Corpus: 268 attributs scripturaux (Dees 1980)  
Indice de similarité (SEM\_{jk}): 100 – Distance Euclidienne Moyenne (DEM\_{jk})

Carte 33: Corpus: 222 vecteurs de localisation (Dees 1983)  
Indice de similarité (SEM\_{jk}): 100 – Distance Euclidienne Moyenne (DEM\_{jk})

Hachures blanches: scores minimal et maximal

Cartes 33-35: Cartes choroplèthes de la synopse de 87 ou 85 valeurs maximales issues des mensurations de similarité précédentes



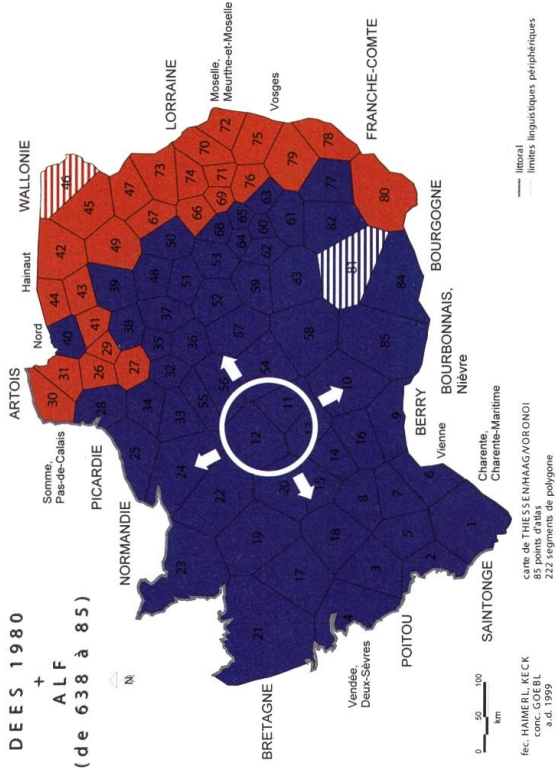


Carte 39: Corpus: 170 attributs scripturaux *vocaliques* (Dees 1980)  
 Indice de similarité (SEM<sub>jk</sub>): 100 – Distance Euclidienne Moyenne (DEM<sub>jk</sub>)

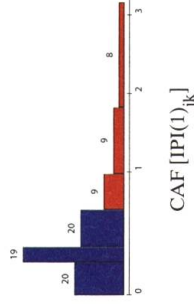
Carte 40: Corpus: 396 cartes de travail (de nature *vocalique*) tirées de l'ALF  
 Indice de similarité: Indice Pondéré d'Identité (avec le poids 1) [IPI(1)<sub>jk</sub>]

Hachures blanches: scores minimal et maximal

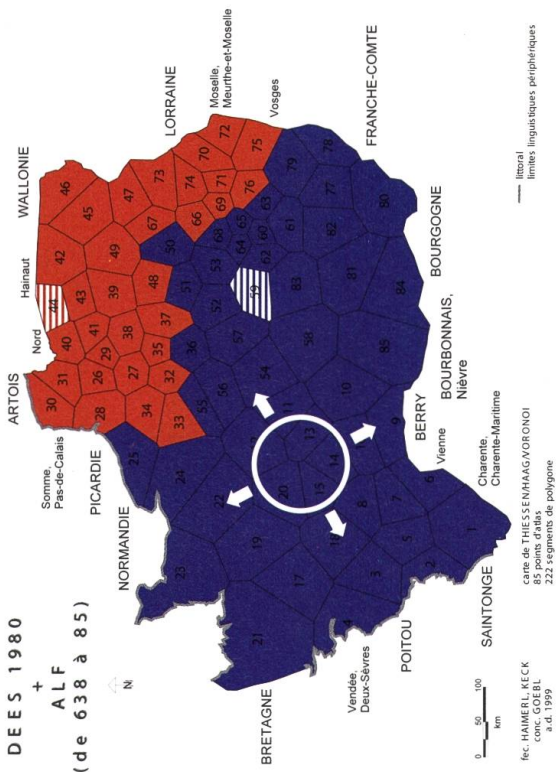
Cartes 39-40: Cartes choroplèthes de la synopse de 87 ou 85 coefficients d'asymétrie de Fisher (CAF) issus des mensurations de similarité précédentes



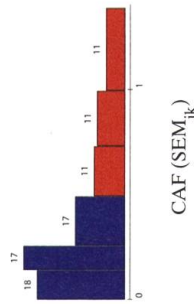
Distribution de fréquence (CAF)  
MEDMW 6-tuple



Algorithme d'intervallisation  
MEDMW 2-tuple



Distribution de fréquence (CAF)  
MEDMW 6-tuple



Algorithme d'intervallisation  
MEDMW 2-tuple

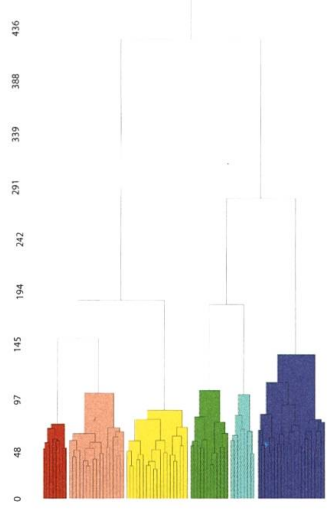
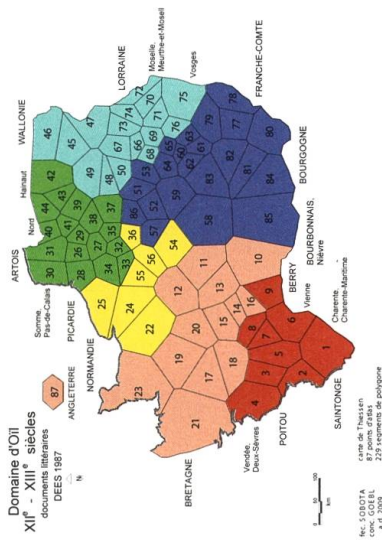
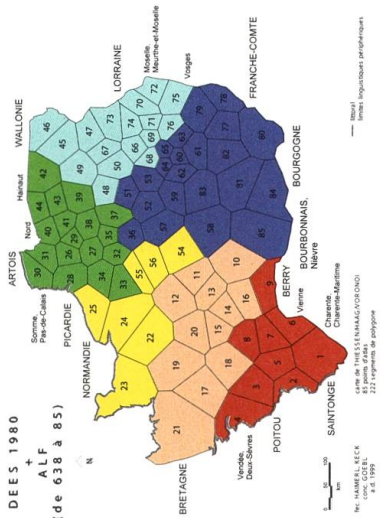
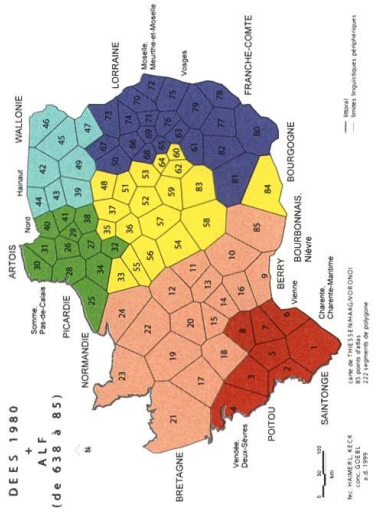


Carte 42: Corpus: 316 cartes de travail (de nature consonantique) tirées de l'ALF  
Indice de similarité: Indice Pondéré d'Identité (avec le poids 1) [IPI(1)]<sub>jk</sub>

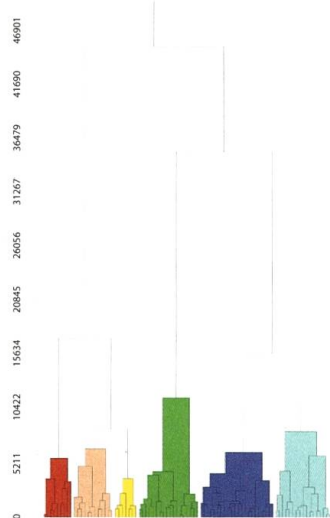
Carte 41: Corpus: 78 attributs scripturaux consonantiques (Dees 1980)  
Indice de similarité (SEM)<sub>jk</sub>: 100 – Distance Euclidienne Moyenne (DEM)<sub>jk</sub>

Hachures blanches : scores minimal et maximal

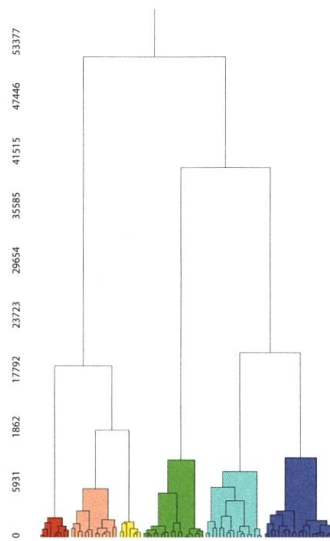
Cartes 41-42: Cartes choroplèthes de la synopse de 87 ou 85 coefficients d'asymétrie de Fisher (CAF) issus des mensurations de similarité précédentes



Carte 45: Corpus: 1 279 cartes de travail (de toutes les catégories linguistiques) tirées de l'ALF  
Indice de similarité: Indice Pondéré d'Identité (avec le poids 1) [IP(1)<sub>jk</sub>]



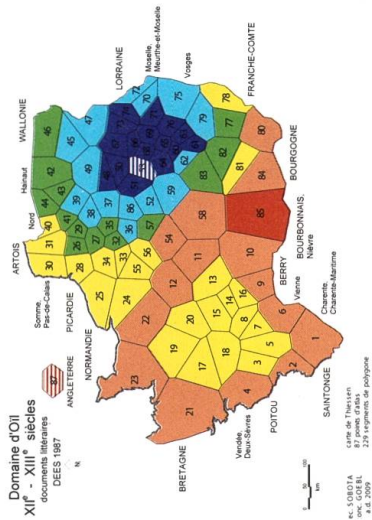
Carte 44: Corpus: 268 attributs scripturaux (Dees 1980)  
Indice de similarité (SEM<sub>jk</sub>): 100 – Distance Euclidienne Moyenne (DEM<sub>jk</sub>)



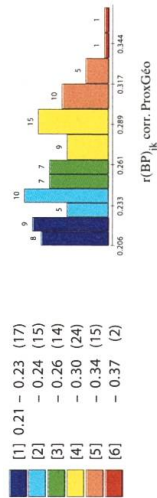
Carte 43: Corpus: 222 vecteurs de localisation (Dees 1983)  
Indice de similarité (SEM<sub>jk</sub>): 100 – Distance Euclidienne Moyenne (DEM<sub>jk</sub>)

Algorithme dendrographique : selon J. Ward, Jr. Classification en six dendrèmes et chorèmes

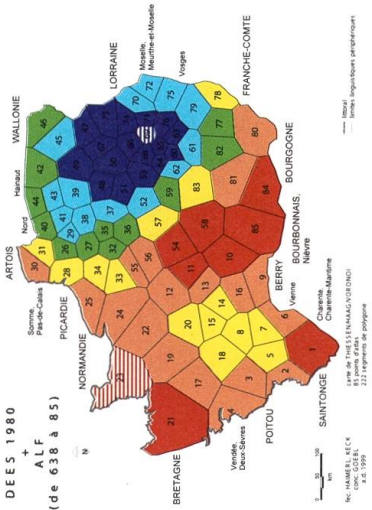
Cartes 43-45: Classification ascendante hiérarchique (CAH) de 87 ou 85 points (de Dees 1983, 1980 et de l'ALF)



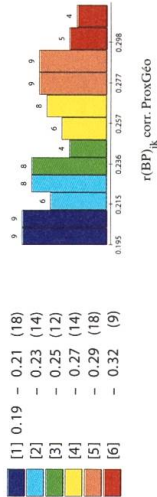
Algorithme d'intervallisation MINMWMAX 6-tuple



Carte 46: matrices carrées corrélées, issues des corpora suivants:  
 a) 222 vecteurs de localisation (Dees 1983)  
 Indice de similarité: coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]$   
 b) proximités géographiques (ProxGéo = 1 - d) selon le théorème de Pythagore (d)

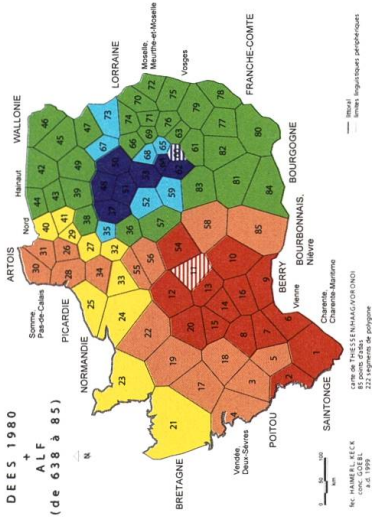


Algorithme d'intervallisation MINMWMAX 6-tuple

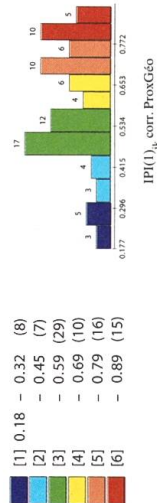


Carte 47: matrices carrées corrélées, issues des corpora suivants:  
 a) 268 attributs scripturaux (Dees 1980)  
 Indice de similarité: coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]$   
 b) proximités géographiques (ProxGéo = 1 - d) selon le théorème de Pythagore (d)

Hachures blanches : scores minimal et maximal



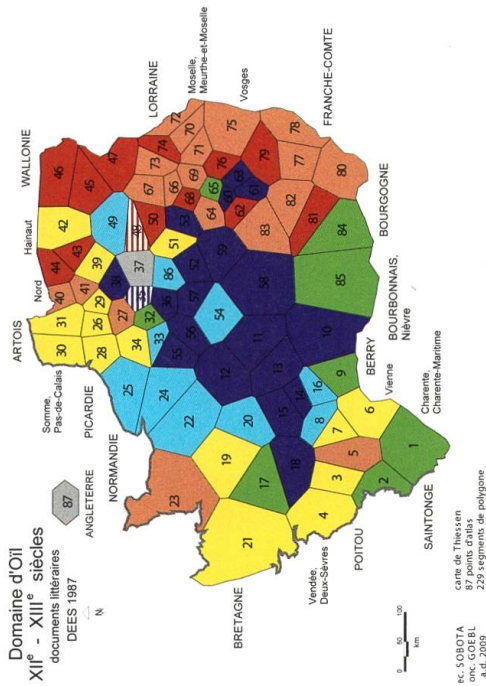
Algorithme d'intervallisation MINMWMAX 6-tuple



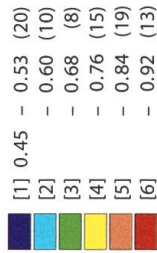
Carte 48: matrices carrées corrélées, issues des corpora suivants:  
 a) 1279 cartes de travail (toutes les catégories linguistiques) tirées de l'ALF  
 Indice de similarité: Indice Pondéré d'Identité (avec le poids 1)  $[IPI(1)_{jk}]$   
 b) proximités géographiques (ProxGéo = 1 - d) selon le théorème de Pythagore (d)

Cartes 46-48: Cartes choroplèthes montrant les con- et divergences diatopiques entre les gestions *linguistique* (scripturale et basilectale) et *euclidienne* de l'espace du domaine d'Oïl

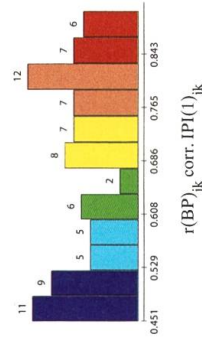




Algorithme d'intervallisation  
MINMWMAX 6-tuple



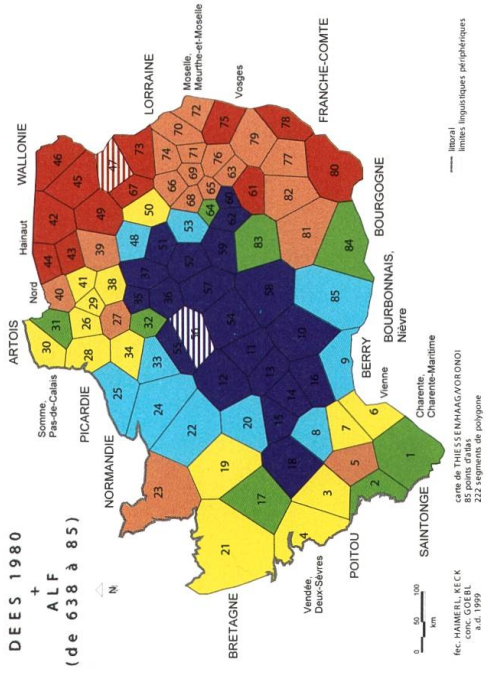
Distribution de corrélation  
MINMWMAX 12-tuple



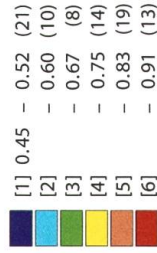
Carte 49: matrices carrées corrélées, issues des corpora suivants:

- a) 222 vecteurs de localisation (Dees 1983)
- Indice de similarité: coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]$
- b) 1 279 cartes de travail (toutes les catégories linguistiques) tirées de l'ALF
- Indice de similarité: Indice Pondéré d'Identité (avec le poids 1)  $[IPI(1)_{jk}]$

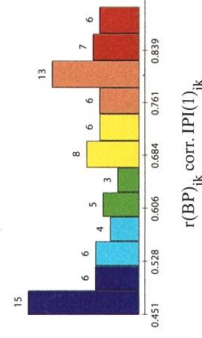
Hachures blanches: scores minimal et maximal



Algorithme d'intervallisation  
MINMWMAX 6-tuple



Distribution de corrélation  
MINMWMAX 12-tuple



Carte 50: matrices carrées corrélées, issues des corpora suivants:

- a) 268 attributs scripturaux (Dees 1980)
- Indice de similarité: coefficient de corrélation de Bravais et Pearson  $[r(BP)_{jk}]$
- b) 1 279 cartes de travail (toutes les catégories linguistiques) tirées de l'ALF
- Indice de similarité: Indice Pondéré d'Identité (avec le poids 1)  $[IPI(1)_{jk}]$

Cartes 49-50: Cartes choroplèthes montrant les con- et divergences diatopiques entre les gestions *scripturale* (littéraire et non littéraire) et *basilectale* de l'espace du domaine d'Oïl