Dialectologia 1 (2008), 3-43.

SUR LE CHANGEMENT MACROLINGUISTIQUE SURVENU ENTRE 1300 ET 1900 DANS LE DOMAINE D'OÏL. UNE ÉTUDE DIACHRONIQUE D'INSPIRATION DIALECTOMÉTRIQUE[1]

Hans Goebl

Universität Salzburg

Hans.Goebl@sbg.ac.at

Abstract

With the supply of 4 figures and 24 closely interpreted dialectometrical maps, this paper analyses the linguistic change of the geolinguistic deep structures in Northern France (Domaine d'Oïl) between 1300 and 1900. As a matter of fact, the result will show – with some well explainable exceptions – the great stability of these deep structures.

The two empirical bases are, on the one hand, the scripta-atlas of the Amsterdam Romanist Anthonij Dees of Northern France (published in 1980) and the "Atlas linguistique de la France" (ALF) of Jules Gilliéron and Edmond Edmont (1902-1910). Whereas the Dees-atlas is based on the philological exploitation of 3300 medieval non literary documents (charters) and therefore refers to written materials, the ALF contains oral materials which have been collected by field work and standardized interviews with competent dialect speakers.

For to comparison of the two empirical levels (*scripturality* and *orality*) we use well experienced methods of the "Salzburg School of Dialectometry" like similarity maps, parameter maps, interpoint maps, correlation maps and trees.

Key words

Scripta, scriptologie, domaine d'Oïl, dialectologie gallo-romane, dialectométrie, Moyen Age, changement linguistique, diachronie.

1. Remarques préliminaires

La finalité de la présente étude est essentiellement d'ordre diachronique. Il s'agit d'évaluer l'ampleur des modifications survenues entre 1300 et 1900, et partant dans un laps de temps de six siècles, sur le plan de l'aménagement linguistique[2] de l'espace du domaine d'Oïl à proprement parler (c'est-à-dire à l'exclusion du domaine francoprovençal). Evidemment, la précision de nos constatations en la matière dépend avant tout de la qualité

[1] Le présent article constitue une version revue et corrigée d'une contribution antérieure parue sous le même titre dans la revue slovène *Linguistica* (46, 2006, 3-43).

[2] Il s'agit d'un terme technique que nous utilisons depuis un certain temps. Celui-ci sert à désigner l'action sémiotique qu'exercent les humains sur l'espace par l'utilisation conjointe de leur faculté langagière pour la *communication* d'un côté et la *démarcation* de l'autre. Ce faisant, nous reprenons une idée chère à Jean Séguy (cf. Séguy 1973, passim).

©Universitat de Barcelona

des sources empiriques à notre disposition, à savoir de l'atlas scripturaire du regretté Anthonij Dees (1928-2001) d'un côté (valable, toute somme faite, pour le dernier tiers du 13^{e} siècle), et de l'ALF de Jules Gilliéron (1854-1926) et d'Edmond Edmont (1849-1926) de l'autre, dont les données ont été collectées entre 1897 et 1901. Bien que les deux sources utilisées par nous, disposent de la même structuration formelle en N points d'atlas et p planches (ou cartes) d'atlas – qui est d'ailleurs celle de tous les atlas linguistiques – il se creuse néanmoins, entre elles, une différence tout autre qu'insignifiante, à savoir celle qui existe entre les plans de l'écrit (Dees) et de l'oral (ALF).

Il faut y ajouter la question de la fiabilité documentaire des matériaux (éditions de textes non littéraires) dépouillés par A. Dees et son équipe. Comme, dans cette étude, il s'agit aussi d'une comparaison basée sur des données de masse et, par conséquent, sujette à la « loi des grands nombres », il est également indiqué de se demander dans quelle mesure, dans la masse des données analysées par A. Dees, le nombre des transcriptions correctes (surtout quand il s'agit de graphies micro-régionales ou carrément locales) dépasse celui des erreurs de lecture ou de transcription. Cependant, les résultats de l'analyse globale des données de l'atlas de Dees sont plutôt rassurants et suggèrent que leur fiabilité documentaire (ou empirique) est suffisamment grande pour une analyse quantitative globalisante. Ceci dit, il va de soi que, pour une analyse qualitative et philologique au sens strict du terme, les données de l'atlas de Dees pourraient se présenter sous un autre jour: cf. à ce sujet Völker 2003, 57-66.

Mis à part son caractère éminemment diachronique, cette étude représente aussi une contribution à la *linguistique quantitative* et, plus particulièrement, à la *dialectométrie*. Ce sont là deux domaines fortement imprégnés de rigueur méthodique et de précision terminologique. C'est pourquoi la lecture des pages qui suivent ne sera pas facile et leur compréhension immédiate ne sera pas garantie d'entrée de jeu. Nous espérons pourtant que la qualité iconique des 4 figures et des 24 cartes à couleur (toutes en annexe) rendra la tâche de lecture et de compréhension moins ardue à nos lecteurs.

1.1. *L'« Atlas des formes et des constructions des chartes françaises du 13^{e} siècle » d'Anthonij Dees (1980)*

L'atlas de Dees repose sur le dépouillement de quelque 3300 chartes oïliques du 13^{e} siècle, toutes publiées et dont la localisation a pu se faire avec suffisamment de précision.

©Universitat de Barcelona

Elles appartiennent, du point de vue textologique, à la catégorie des documents *non littéraires* (cf. Dees 1980, X)[3]. Quant à leur provenance régionale (ou « localisation »), Dees les a répertoriées dans 85 unités spatiales dont 28 figurent aussi sur les 282 planches de l'atlas publié. Ces 28 unités spatiales peuvent être considérées comme des centres scripturaires «majeurs» (voir la colonne *Macro (Ma)* sur la figure 1) alors que les autres en constituent des subdivisions diatopiques et ne représentent, de ce fait, que des centres scripturaires «mineurs» (voir la colonne *Micro (Mi)* sur la figure 1): cf. Dees 1980, 307-312 et la liste respective présentée ici en appendice (figure 1).

Les 3300 chartes prises en considération ont été analysées à l'aide d'un questionnaire phonético-graphique et morphologique dont les items correspondent exactement aux en-têtes des 282 cartes de l'ouvrage publié (Dees 1980). On en trouve une liste détaillée aussi à la fin de l'atlas de Dees (« Inventaire des formes trouvées », 322-369). Une autre liste (« Formes et constructions étudiées », ibidem 313-321) renseigne sur la structuration interne de l'atlas dédié, aux termes de son titre, à l'étude « des formes et des constructions » de la scripta non littéraire du 13^{e} siècle.

Pour l'analyse de la variabilité graphique bien connue des documents médiévaux, Dees a décidé de saisir, pour chaque item, les «alternances» entre deux groupes de graphies (ou de constructions morphologiques) concurrents. Evidemment, la définition de ces « alternances » ne saurait jamais être absolue et dépend de certaines présuppositions individuelles. Ceci implique qu'un autre chercheur, dans une situation analytique similaire, aurait pu adopter une solution différente[4]. En voici un exemple relatif à la carte 7 de l'atlas de Dees. Il s'agit du pronom personnel «moi» et de l'alternance entre la forme *mi* d'un côté et les formes *mei*, *mey*, *moi* et *moy* de l'autre. Les formes en *mi* constituent, dans l'atlas de Dees, le groupe 1 alors que les autres formes sont rangées dans le groupe 2. Seuls les scores du groupe 1, moyennant six paliers à hachures en noir-et-blanc, sont mis en carte alors qu'une mise en carte similaire n'a pas été faite pour les scores du groupe 2: voir aussi la figure 2 (en annexe).

[3] Signalons encore l'existence d'un autre atlas scripturaire de Dees, publié en 1987 et relatif aux textes littéraires des 12^{e} et 13^{e} siècles, dont, malheureusement, les fichiers électroniques originaux utilisés au cours de la publication, ne subsistent plus. Sa valeur documentaire a été mise en doute par beaucoup de spécialistes de la tradition textologique de la littérature médiévale française: cf. à ce sujet le rapport détaillé de Völker 2003, 57-66. Néanmoins, je persiste à croire qu'une analyse dialectométrique de cet atlas, comparable à celle que nous présentons dans cet article, sera également fort intéressante: cf. aussi Goebl 1993.

[4], Dees est d'ailleurs conscient de ce fait lui-même. Quant aux variantes « dichotomisées » de l'article contracté « du », il s'exprime de la manière suivante: « D'autres dichotomisations de la même série des variantes de "du" auraient pu être envisagées; en général la dichotomie qui a été finalement retenue (parfois après de longs tâtonnements) est celle qui donne le meilleur résultat cartographique en termes de cohérence des aires et qui permettra souvent de se faire une idée de l'histoire des formes étudiées. » (Dees 1980, XI).

©Universitat de Barcelona

Quant au bilan numérique des deux groupes en question, l'on en trouve, en bas des 282 cartes de l'atlas de Dees, les listes avec les scores respectifs (toujours en chiffres absolus et relatifs), si bien que des lecteurs assoiffés de statistiques peuvent contrôler la correction de tous les calculs effectués et aussi de la mise en carte des graphies appartenant au groupe 1.

En ce qui concerne la mensuration de la variation scripturaire inhérente à son corpus, Dees a choisi de retenir, pour les chartes de chacun des 85 centres scripturaires considérés, le pourcentage de celles qui disposent de graphies appartenant ou bien au groupe 1 ou bien au groupe 2[5]. En voici un exemple de calcul relatif à la carte 1, dédiée au pronom personnel « je », et au centre scripturaire (majeur) 01 (Charente, Charente-Maritime). La liste sur la «provenance des chartes» (Dees 1980, 307-312) nous dit que cette macro-unité territoriale dispose d'un effectif total de 118 chartes. La moitié de gauche de la légende numérique au pied de la carte 1, nous renseigne que, de ces 118 chartes, 111 peuvent être utilisées selon la logique de l'« alternance » sur laquelle repose la carte 1:

groupe 1 (visualisé): *g*, *ge*, *gie*, etc.

groupe 2 (non visualisé): *i*, *ie*, *iou*, *je*, *jo*, etc.

Or, l'introspection de ces 111 chartes a montré que l'écrasante majorité – à savoir 105,24[6] – fournissent des graphies en *g*, *ge*, *gie*, etc. (groupe 1) alors que le reste – c'est-à-dire 5,76 chartes – contiennent des graphies retenues comme «alternatives» (ou complémentaires) en *i*, *ie*, *iou*, *je*, *jo*, etc.

La mise en pourcentage successive fournit les résultats suivants: 105,24 : 111 = 94,81% et 5,76 : 111= 5,19%. On trouve ces valeurs dans la moitié de droite de la légende numérique de la carte 1. Or, ce sont elles que nous avons utilisées pour l'établissement de notre matrice de données dialectométrique (voir la figure 2).

Du point de vue quantitatif, l'atlas de Dees constitue donc, tel qu'il se présente sous sa forme publiée, une matrice bidimensionnelle avec 28 entités territoriales et 282 attributs scripturaires dont les scores occupent le niveau de l'échelle métrique. Cet agencement formel

[5] « Ce problème a été résolu dans ce sens que nous avons systématiquement ramené les variations à étudier à des dichotomies. Cette dichotomisation, qui consiste à répartir l'ensemble des formes rencontrées sur deux classes, dont la première est caractérisée par une ou plusieurs marques que la deuxième ne comporte pas, peut être illustrée par l'exemple relativement simple des variantes de "tous", comprenant entre autres les graphies *tous*, *touz*, *tos*, *toz*. Ces quatre graphies se ramènent à deux alternances, à savoir *-ou-*: -o- et *-s-* : *-z-*. » (Dees 1980, XI).

[6] Précisons que les effectifs non entiers (105,24 et non pas 105 ou 106) s'expliquent par la présence conjointe, dans une charte donnée, de graphies appartenant aux deux groupes mis en contraste: « Si la charte comporte des variantes appartenant à l'une et à l'autre de deux classes opposées, le témoignage est divisé proportionellement aux fréquences absolues des deux variantes dans ce document. » (Dees 1980, XII).

©Universitat de Barcelona

le prédestine, en quelque sorte, à être dépouillé par voie dialectométrique[7]. Ajoutons que la matrice de données utilisée dans l'atelier informatique d'A. Dees à Amsterdam (et dont la direction était assurée, jadis, par Pieter van Reenen), disposait, du côté des entités territoriales, non seulement de 28, mais bien plutôt de 85 unités spatiales qui, évidemment, couvraient l'espace du domaine d'Oïl d'une façon beaucoup plus détaillée; voir à ce sujet, une fois de plus, la figure 1 (en annexe).

Ce n'est qu'en 1996 que la coïncidence de deux événements propices a fini par dégager la voie pour une véritable « dialectométrisation » des données de l'atlas de Dees:

1) la remise des fichiers électroniques de l'atlas de Dees par Pieter van Reenen à l'auteur de cet article,

2) l'arrivée d'un nouveau collaborateur dialectométrique à Salzbourg, Guillaume Schiltz[8].

Evidemment, nos activités dialectométriques ne se sont pas arrêtées là et ont continué à se développer dans deux directions:

a) à partir de 1998, par la création d'un nouveau logiciel dialectométrique ultra-puissant appelé VDM ("Visual DialectoMetry") de la part d'Edgar Haimerl,

b) à partir de 1997, par la dialectométrisation de presque la moitié des 1421 planches originales de l'ALF.

Les analyses qui suivent, reposent donc, du point de vue informatique (VDM) et documentaire (disponibilité électronique conjointe des données de l'atlas de Dees et de l'ALF), sur des antécédents plutôt complexes et échelonnés sur un laps de temps relativement long (1996-2005).

Nous avons tiré, des fichiers électroniques qui nous ont été remis par Piet van Reenen en 1996, deux matrices de données (métriques) qui diffèrent, l'une de l'autre, le long de l'axe du nombre des attributs scripturaires. Alors que la première – dans la mesure du possible – est exempte de données manquantes, l'autre embrasse la totalité des 268 cartes qui nous ont été confiées. Pour une raison qui nous échappe, ces données disposent d'un total de seulement 268 cartes (ou attributs scripturaires) alors que l'atlas de Dees (1980) comprend 298 attributs scripturaires.

Le tableau suivant renseigne sur les dimensions des deux matrices de données (métriques) mentionnées.

[7] Dans notre compte rendu de l'atlas de Dees, nous avons déjà évoqué l'utilité d'une dialectométrisation de ce dernier (Goebl 1982, 283).
[8] Cf. les deux contributions (Goebl/Schiltz 2001 et de Schiltz 1997), issues de cette coopération.

©Universitat de Barcelona

Catégorie linguistique	sigle	Corpus réduit (RED): nombre des cartes (attributs scripturaires) disponibles	Corpus intégral (TOT): nombre des cartes (attributs scripturaires) disponibles
vocalisme	VOC	119	170
consonantisme	CONS	66	78
morphologie	MORPH	18	20
∑		203 (RED)	268 (TOT)

1.2. *L'« Atlas linguistique de la France » (ALF)*

Cette étude, d'inspiration comparative par excellence, se compose de deux volets: le premier se réfère à l'atlas de Dees et le second à l'ALF. Evidemment, nous entendons par «ALF» aussi la version dialectométrisée de l'ouvrage original telle qu'elle figure dans la matrice de données issues des travaux de «taxation» qui se sont déroulés à Salzbourg entre 1997 et 2000. La dialectométrisation de l'ALF – dont le premier bilan a eté publié en 2000 (cf. Goebl 2000) – avait comme but d'analyser (ou de «taxer») un nombre aussi grand que possible de cartes originales de l'ALF appartenant à différentes catégories linguistiques (lexique, vocalisme, consonantisme etc., avec leurs sous-catégories respectives).

Le travail de taxation avait été réalisé par une équipe de plusieurs jeunes romanistes salzbourgeoises qui se sont dédiées à cette tâche méticuleuse (et aussi harassante) pendant trois ans (1997-2000)[9]. C'est parallèlement à la saisie taxatoire des données de l'ALF que M. Haimerl a développé le logiciel VDM d'où la possibilité d'appliquer aux données taxées – aussitôt qu'elles ont été enfournées dans l'ordinateur – la gamme entière des calculs dialectométriques désormais facilement accessibles[10].

Pour une comparaison directe des données de Dees et de l'ALF, il fallait encore procéder à la réduction de la densité du maillage de l'ALF. Ce travail, ainsi que la confection de tous les graphiques de cet article, a été réalisé par Slawomir Sobota, un autre collaborateur fidèle et compétent du laboratoire dialectométrique de Salzbourg.

La structuration interne[11] des matériaux dialectométrisés de l'ALF tels qu'ils ont été utilisés dans cette étude, se présente donc de la façon suivante:

[9] Il s'agit de Barbara Aigner, Irmgard Dautermann, Hildegund Eder, Susanne Oleinek et Annette Ida Schatzmann. Qu'elles reçoivent ici l'expression de notre profonde reconnaissance!

[10] Pour d'autres détails (techniques et logistiques) relatifs à la dialectométrisation de l'ALF, cf. aussi Goebl 2002, 67 et 2003, 61-62.

[11] Nous ne parlons ici que des catégories linguistiques prises en considération dans cette étude.

©Universitat de Barcelona

catégorie linguistique	sigle	Nombre des cartes-ALF originales dépouillées	Nombre des «cartes de travail»[12] tirées des cartes-ALF originales
lexique	LEX	463	471
vocalisme	VOC	242	612
consonantisme	CONS	227	505
∑			1687 (TOT)

La masse considérable de 1687 cartes de travail disponibles – issues de l'analyse taxatoire de 626 planches originales tirées par voie aléatoire de l'effectif total des 1421 cartes de l'ALF – confère à nos analyses dialectométriques de l'ALF une très grande fiabilité statistique. Ceci vaut également pour les sous-catégories du *vocalisme* et du *consonantisme*. En ce qui concerne la fiabilité statistique des données de l'atlas de Dees par contre, la situation n'est pas aussi rassurante, certes, mais encore suffisamment solide pour permettre l'analyse comparative envisagée.

1.3. *Conventions (carto)graphiques*

Pour une bonne intelligence du message iconique (et aussi quantitatif) des 24 cartes de cet article, les lecteurs sont invités à se familiariser, d'entrée de jeu, avec les conventions (carto)graphiques utilisées par nous.

1.3.1. Préparation cartographique des réseaux de Dees et de l'ALF

Pour les deux projets dialectométriques (Dees et ALF) en question, il y a un réseau unique. Celui-ci dispose de 85 unités spatiales (appelées aussi « points »). Quant aux points de l'atlas de Dees, on en trouve une liste dans la figure 1 (voir la colonne *Micro (Mi)*). Pour ce qui est du réseau de l'ALF par contre, nous avons dû l'amputer de ses parties non françaises et non oïliques d'abord et l'élaguer convenablement par la suite. L'étape suivante consistait dans l'application de la géométrie de Delaunay-Voronoï (triangulation, construction des médiatrices des côtés de triangles respectifs, fusion de ces dernières et construction des

[12] Nous entendons par « carte de travail » le résultat d'une analyse taxatoire appliquée à une planche originale de l'ALF. Du point de vue formel, chaque carte de travail correspond à un vecteur d'attribut (horizontal) de la matrice de donnees (voir aussi la figure 4).

©Universitat de Barcelona

polygones de Voronoï (appelés aussi polygones de «Thiessen» ou de «Haag»)[13]. L'utilité du réseau polygonal ainsi constitué est double: il permet la confection tant de cartes choroplèthes (c'est-à-dire à plages coloriées) que de cartes isarithmiques (c'est-à-dire munies de côtés de polygone à épaisseur et coloriage variables).

1.3.2. Mise en carte du bilan numérique des calculs dialectométriques

Evidemment, la dialectométrie n'a pas pour but de se gargariser de chiffres, mais vise en premier lieu à les visualisier en bonne et due forme pour les faire parler ensuite aux yeux des linguistes. Pour ce faire, elle se sert, depuis longtemps, de tous les gadgets de la cartographie statistique internationale parmi lesquels figurent, entre autres:

a) l'utilisation de la gamme (et de l'agencement naturel) des couleurs de l'arc-en-ciel (ou du spectre solaire) pour la visualisation claire et distincte de la variabilité numérique discutée,

b) l'usage de 2 à 8 paliers chromatiques pour le rendement iconique de la variabilité numérique considérée,

c) le recours à différents algorithmes d'intervallisation (ici: MINMWMAX[14] et MEDMW[15]) pour la constitution « objective » de classes (ou paliers) iconiques différentes,

d) la construction d'histogrammes (ici: toujours à droite, en bas de page) pour la visualisation adéquate des propriétés statistiques des différentes distributions de fréquence considérées[16].

Le choix du nombre des paliers iconiques et de l'algorithme de visualisation (facilement réalisable à l'aide du logiciel VDM) se fait toujours en fonction de la finalité exploratoire de l'analyse en question. Les modalités de la visualisation doivent donc toujours rester au service de la démarche scientifique du linguiste et l'aider dans le développement et la vérification de ses vues théoriques sur le sujet de recherche en question.

---

[13] Pour une présentation détaillée de ces procédures, cf. Goebl 1981, 363-365 et 1984 I, 90-92.

[14] Les scores à visualiser sont d'abord répartis des deux côtés de la moyenne arithmétique. Ensuite les écarts entre le minimum et la moyenne arithmétique et le maximum et la moyenne sont divisés – pour une visualisation en six classes – par trois. C'est à l'aide des valeurs ainsi déterminées qu'il est possible de définir, toujours des deux côtés de la moyenne arithmétique en question, les seuils numériques (inférieurs et supérieurs) des six classes respectives.

[15] Les scores à visualiser sont répartis des deux côtés de la moyenne arithmétique d'abord et répartis par la suite, séparément des deux côtés de la moyenne arithmétique, dans des classes égales en nombre d'objets (éléments, points d'atlas etc.).

[16] Pour plus de détails cf. Goebl 1981, 363 et 1984 I, 97-98. Les courbes gaussiennes (dites aussi «normales») qui recouvrent les histogrammes, ont été calculées à partir de la moyenne arithmétique et de l'écart-type de la distribution de fréquence respective, et permettent d'en contrôler rapidement certaines propriétés statistiques (telle la symétrie, entre autres).

©Universitat de Barcelona

1.4. De la matrice de données à la matrice de similarité (ou de distance)

Toute recherche dialectométrique est organisée sous la forme d'une chaîne dont les différents maillons constituent autant de «rond-points» méthodiques. Evidemment, la chaîne dialectométrique à proprement parler commence par l'établissement de la matrice de données. Les données de cette dernière se situent, quant à l'atlas de Dees, au niveau de l'échelle *métrique* alors que les données de la matrice de données de l'ALF occupent le niveau de l'échelle *nominale* (ou *qualitative*): voir la moitié de gauche de la figure 4. Cette différence d'ordre métrologique influe directement sur le choix des indices de similarité (de corrélation ou de distance) dont on a besoin pour calculer la similarité (linguistique) réciproque entre les 85 «points» scripturaires (Dees) ou dialectaux (ALF). Or, les manuels de taxométrie modernes[17] en offrent une panoplie très nourrie dont nous nous sommes servis dès le début de nos travaux dialectométriques dans les années 70 du siècle dernier.

Pour l'analyse dialectométrique des données de l'atlas de Dees, nous avons utilisé les trois indices suivants:

a) le coefficient de corrélation (de Bravais et Pearson) [$r(BP)_{jk}$][18],

b) la «Similarité Euclidienne Moyenne» (SEM_{jk}): voir la figure 3,

c) la «Distance Euclidienne Moyenne» (DEM_{jk})[19]: voir la figure 3.

La relation entre SEM_{jk} et DEM_{jk} se définit comme suit: $SEM_{jk} + DEM_{jk} = 100$.

Pour ce qui est des données de l'ALF par contre, dont le niveau métrologique est *nominal* (ou *qualitatif*), nous nous sommes servis de trois indices couramment utilisés en matière de dialectométrie, à savoir:

a) de l'«Indice Relatif d'Identité» (IRI_{jk})[20] qui, à cause de son usage fréquent, constitue, en quelque sorte, l'indice de similarité standard de la dialectométrie, puis

b) de l'«Indice Relatif de Distance» (IRD_{jk}) dont la relation face à l'IRI_{jk} est régie par la formule suivante: $IRD_{jk} + IRI_{jk} = 100$, et, finalement,

c) de l'«Indice Pondéré d'Identité (avec le poids 1)» ($IPI(1)_{jk}$)[21] qui permet de capter mieux que ne le pourrait faire l'IRI_{jk}, des structures diatopiques «plates» enfouies dans la profondeur des données à analyser.

---

[17] Cf. Bock 1974 (en allemand), Chandon/Pinson 1981 (en français) et Sneath/Sokal 1973 (en anglais).
[18] Cf. Chandon/Pinson 1981, 65 ou Sneath/Sokal 1973, 137-140.
[19] Pour une description détaillée de la DEM_{jk} cf. Bock 1974, 36-39, Chandon/Pinson 1981, 58-59 et Sneath/Sokal 1973, 121-125.
[20] Cf. Goebl 1981, 357-361 et 1984 I, 75-78.

©Universitat de Barcelona

Pour la sélection d'un de ces indices, il n'y a pas de critères fixes ou prétablis. Leur finalité est comparable à celle d'un outil dont on se sert pour achever tel ou tel ouvrage. Comme, en matière de taxométrie (et, partant, aussi de dialectométrie), il s'agit de *découvrir* (et non pas d'*inventer* ou d'y *projeter*!), par voie quantitative, des *structures cachées* au fond des données analysées, l'utilité et le rendement taxométriques de tel indice s'avèrent avant tout à travers sa capacité de détecter, avec suffisamment de clarté, de telles structures de profondeur. Ceci signifie que, pour ce faire, le dialectométricien doit se livrer à un certain nombre d'expériences qui, elles, sont d'ailleurs facilement réalisables à l'aide du logiciel VDM. C'est que celui-ci dispose, dans sa charpente méthodique, d'un grand nombre d'indices de similarité et de distance. Ceci n'empêche que, pour des raisons purement formelles (ou mathématiques), la nature métrologique des données à analyser (susceptibles d'occuper, alternativement, les échelles *nominale*, *ordinale* ou *métrique*) impose le choix de certains indices et en déconseille l'emploi d'autres.

2. Six coups d'oeil comparatifs sur les structures de profondeur[22] de l'atlas de Dees et de l'ALF

Nous procéderons par la suite, à titre d'exemple, à six comparaisons (ou « plateformes ») typodiagnostiques dont l'orientation méthodique et la finalité gnoséologique seront très variées. Du point de vue formel, il s'agit d'un côté de comparaisons purement visuelles (basées, bien sûr, sur des prérequis formels et statistiques strictement identiques), et de corrélations spatiales au sens strict du terme de l'autre (transposées, elles aussi, en visualisations adéquates). Le but de ces rapprochements comparatifs est de permettre aux lecteurs de juger eux-mêmes des étonnantes ressemblances (mais aussi de certaines dissemblances non négligeables) qui existent entre l'aménagement géolinguistique du domaine d'Oïl à la fin des 13^{e} et 19^{e} siècles[23]. Chacune des 12 pages de l'annexe où figurent

[21] Cf. Goebl 1987, 70-79.

[22] La considération quantitative de données de masse présuppose qu'il y ait, à l'intérieur des données analysées, un grand nombre de structurations non aléatoires et hiérarchisées entre elles. L'analyse quantitative permet de les découvrir, quitte d'ailleurs à opérer, suivant le cas, dans différents secteurs (ou «étages») des données à analyser. Il en résulte la nécessité de différencier les structures détectées suivant le *degré de leur insertion* dans la totalité des matériaux analysés: d'où la distinction entre structures de «profondeur» et de «superficie».

[23] La précision des analyses dialectométriques et la perfection des graphiques présentées dans cet article dépassent celles de nos contributions antérieures (Goebl 1998 et 2001 ainsi que Goebl/Schiltz 2001). Ceci est dû avant tout au logiciel VDM et à ses capacités numériques et visualisatrices.

©Universitat de Barcelona

les 24 cartes illustratives, sera arrangée de façon identique: le réseau de l'atlas de Dees y figure toujours à gauche, celui de l'ALF à droite.

2.1. *Plateforme comparative 1: la mensuration de la similarité*

Voir les cartes 1-8.

L'outil dialectométrique de la mensuration de similarité permet de saisir les rapports de similarité (ressemblance, identité etc.) linguistiques qui existent entre les N éléments (dialectes, scriptae etc.) d'un réseau géolinguistique donné. Pour un réseau d'atlas de N points et p attributs, il est possible de dresser, au cours de la mensuration de similarité appliquée aux N vecteurs d'objets (verticaux)[24] du réseau, une matrice de similarité (carrée) aux dimensions N fois N, dont les vecteurs contiennent l'ensemble de l'information (quantitative) susceptible d'être réunie par le procédé de la mensuration de la similarité en tant que telle.

Or, chacun des N vecteurs de la matrice de similarité dispose de N-1 scores situés au-dessous de la valeur maximale (de 1 ou 100%) et d'un seul score (s_{jj}) toujours équivalent à 1 (ou 100%), issu d'une mensuration réflexive appliquée au point de référence j lui-même. Par la visualisation d'un tel vecteur, l'on obtient une carte de similarité laquelle, par définition, se compose de N-1 plages (ou polygones) réparties dans l'espace et coloriées en fonction de leur rendement numérique, ainsi que d'un point de repère (ou de référence) j laissé toujours en blanc.

La lecture d'une carte de similarité est facile. Elle nous renseigne, par la distribution, dans l'espace étudié, d'un nombre pair de couleurs «chaudes» (p. ex. rouge, orange, jaune) et de couleurs «froides» (p. ex. vert, bleu clair, bleu foncé), sur la «position» du dialecte (parlé ou scripturaire) en question parmi les dialectes restants du réseau examiné. Soit dit entre parenthèses, la carte de similarité constitue le pilier central de la charpente heuristique (et aussi théorique) de la dialectométrie entière.

Pour l'établissement des cartes 1, 3, 5 et 7 (relatives aux données de l'atlas de Dees), nous avons utilisé le coefficient de corrélation de Bravais et Pearson ($r(BP)_{jk}$)[25] alors que les

[24] Se référer, pour une bonne intelligence de ce processus, à la figure 4.

[25] Les valeurs du $r(BP)_{jk}$ oscillent entre -1 (pour une corrélation [linéaire] absolument négative) et +1 (pour une corrélation [linéaire] absolument positive). Précisons que, dans ce contexte, la notion de «corrélation» remplace celle de «similarité (dialectale, scripturaire etc.)» et aussi qu'à la fin de son atlas (1980, 370-371), Dees lui-même s'est servi deux fois du $r(BP)_{jk}$ pour le calcul de la similarité scripturaire entre plusieurs documents médiévaux. La même remarque vaut pour Dees 1984 et pour la contribution de Huber/van Reenen-Stein (1988) où l'on trouve une autre application du $r(BP)_{jk}$ aux données de l'atlas de Dees.

©Universitat de Barcelona

scores des cartes 2, 4, 6 et 8 (relatives à l'ALF) ont été calculés à l'aide de l'Indice Pondéré d'Identité (avec le poids 1) $(IPI(1)_{jk})$[26]. C'est par l'application de ces deux indices particuliers que la comparaison dialectométrique des deux niveaux atlantographiques acquiert le plus de précision et de netteté.

Les points de référence, situés aux quatre coins du domaine d'Oïl (dans le Poitou, en Picardie, en Ile-de-France et dans la Franche-Comté), ont été choisis en fonction de leur diversité géographique. Leur «profil choroplèthe» (appelé aussi «profil de similarité») nous renseigne sur la localisation de – métaphoriquement parlant – leurs «meilleurs amis» (voir les polygones en rouge) et aussi de leurs «antipodes (ou pires ennemis)» (voir les polygones en bleu foncé). A remarquer en outre le décalage bien agencé des polygones en position intermédiaire, c'est-à-dire pourvus des couleurs orange, jaune, vert et bleu clair.

D'une comparaison visuelle des quatre paires de cartes, il ressort avec netteté que les aménagements linguistiques de l'espace du domaine d'Oïl tels qu'ils se présentent sur les quatre cartes de similarité scripturaires et leur pendants relatifs aux dialectes modernes, ont dû être sensiblement très proches les uns des autres en 1300 et en 1900. Evidemment, il s'agit d'un bilan global qui repose sur la synthèse quantitative d'un nombre très grand de distributions spatiales particulières (de nature qualitative et quantitative). Néanmoins, cette constatation (ou plutôt découverte) est quelque peu surprenante – pour ne pas dire troublante – surtout à cause du fait qu'elle nous suggère que la structuration de base du domaine d'Oïl, telle que nous la connaissons à partir des données de l'ALF, a déjà dû exister – mutatis mutandis – 600 ans plus tôt[27].

Ajoutons encore – avant de passer au paragraphe suivant – une petite remarque sur la relation qui existe entre la diminution de la similarité linguistique et de la proximité géographique (ou euclidienne) par rapport à un point de référence préalablement déterminé. La relation entre l'aménagement linguistique et euclidien de l'espace est extrêmement complexe. Bien loin d'être uniquement un phénomène linéaire, cette relation connaît des réalisations spatiales très bigarrées. Sur chacune de nos huit cartes de similarité, elle est d'une autre nature et revêt des configurations géographiques très diversifiées. Evidemment, il serait intéressant de l'étudier de plus près comme ceci a été fait par Jean Séguy dans son premier article d'inspiration dialectométrique intitulé « La relation entre la distance spatiale et la distance lexicale », publié en 1971, tout en remplaçant d'ailleurs la recherche de « courbes »,

[26] Pour une explication détaillée de cet indice cf. Goebl 1987, 70-79.
[27] De telles convergences spatio-temporelles ont été constatées, pour des faits linguistiques isolés, déjà à plusieurs reprises: cf. Remacle 1992, 157-169, Goebl 1989, 97-98 et Besch 1985, 1793-1798.

©Universitat de Barcelona

pratiquée jadis par Séguy, par la considération synoptique spatiale des corrélations quantitatives respectives.

2.2. *Plateforme comparative 2: la corrélation spatiale entre les aménagements linguistique et euclidien de l'espace du domaine d'Oïl*

Voir les cartes 9-10 et la figure 4.

A l'aide de la fameuse formule de Pythagore, il est facile de calculer, entre les 85 points de notre réseau combiné (Dees et ALF) et à l'aide des coordonnées cartésiennes (en x et y) y ayant trait, leurs distances (*dist*) ou proximités (*prox*)[28] réciproques. C'est ainsi que l'on obtient une matrice (carrée) de proximité aux dimensions 85 fois 85 où se trouvent stockés les scores de proximité entre tous les 85 points de notre réseau. Or, cette matrice de proximité peut être corrélée, vecteur par vecteur, avec la matrice de similarité (ou de corrélation) linguistique comme le montre la figure 4. Le résultat de ces calculs de corrélation – exécutés, une fois de plus, à l'aide du coefficient de corrélation $r(BP)_{jk}$ – est un ensemble de 85 scores de corrélation qui, eux, doivent être mis en carte par la suite: voir les cartes 9 et 10.

Le profil diatopique qui se dégage de cette mise en synopse est très éloquent. L'on y discerne une nette bipartition du domaine d'Oïl. Alors que l'Est est caractérisé par une très mauvaise coïncidence entre les aménagements linguistique et euclidien de l'espace, l'inverse est vrai pour l'Ouest et le Centre.

Précisons qu'une «mauvaise» coïncidence entre les aménagements linguistique et euclidien de l'espace signifie que, dans la réalité empirique, la similarité linguistique décroît plus vite (ou plus lentement) que ne le suggère le modèle (théorique voire «idéal») de l'évolution diatopique parallèle des deux variables. Or, il est vrai que de telles divergences se manifestent toujours là où de fortes tensions politiques, culturelles ou sociales accélèrent (ou ralentissent) le déploiement diatopique «naturel» de l'activité langagière de l'HOMO LOQUENS. L'inverse est également vrai: ceci signifie que les zones marquées en rouge et orange renvoient à une synchronisation spatiale très marquée entre la configuration euclidienne de l'espace et l'étalage diatopique des similarités interdialectales.

L'agencement iconique convergent des profils choroplèthes des cartes 9 et 10 nous suggère donc que, dans l'optique de notre analyse dialectométrique, les prérequis spatio-

[28] La relation entre distance (*dist*) et proximité (*prox*) est la suivante: *dist* + *prox* = 100.

©Universitat de Barcelona

linguistiques de base du domaine d'Oïl n'ont pratiquement pas bougé entre 1300 et 1900 et que, par conséquent, ils doivent remonter à une époque antérieure à 1300.

2.3. Plateforme comparative 3: les cartes à cloisons (ou: cartes à interpoints en fonction discriminatoire)

Voir les cartes 11 et 12.

Les cartes à cloisons (appelées aussi «cartes à interpoints en fonction discriminatoire») constituent un des outils les plus simples (et fort éloquents à la fois) de la dialectométrie. Du point de vue heuristique, elles correspondent à la méthode bien connue du tracement combiné d'isoglosses. En voici les différentes étapes[29]:

a) Triangulation du réseau (Dees-ALF): sur 85 points, il en résulte 225 côtés de triangle.

b) Construction, pour chacun des 225 côtés de triangle, de la médiatrice respective. Les 225 médiatrices, dûment fusionnées entre elles, finissent par fournir le parquet polygonal du réseau-Dees-ALF.

c) Chacun des 225 côtés de polygone servira de support iconique pour la visualisation non plus de similarités, mais de distances interponctuelles, c'est-à-dire situées entre deux points d'atlas contigus.

d) La visualisation des scores de distance se sert de deux variables iconiques: de l'épaisseur (elle augmente avec la valeur numérique de la distance linguistique interponctuelle) et de la couleur (la couleur bleu foncé est associée au maximum de la distance linguistique interponctuelle, et la couleur rouge au minimum de cette dernière).

C'est ainsi que la syntaxe iconique des deux cartes repose sur le jeu optique différentiel de 225 côtés de polygone d'épaisseur et de couleur variables. Il en résulte un compartimentage alvéolaire spécifique du réseau polygonal qui est très éloquent. L'on y distingue en effet la présence alternative de zones fortement cloisonnées (et sillonnées, de ce fait, de beaucoup de côtés de polygone épais et coloriés en bleu foncé) et d'autres où prédominent les polygones minces et coloriés en rouge, orange et jaune.

Alors que la carte 11 laisse entrevoir, par le biais de l'effet-cloison, quelques-unes des provinces historiques du Nord de la France (Poitou, Maine, Normandie, Artois, Picardie etc.), il n'en est rien pous la carte 12 (relative à l'ALF et, partant, à la date de 1900) qui est

[29] Cf. aussi Goebl 1983 (passim) et 1984 I, 183-196.

©Universitat de Barcelona

caractérisée par une bipartition diatopique très claire: concentration de l'effet-cloison à la périphérie (Sud-Ouest, Nord, Nord-Est, Est) d'un côté, et réduction de l'effet-cloison dans les régions centrales du domaine d'Oïl de l'autre[30]. D'une comparaison visuelle des deux cartes il appert que le changement macrolinguistique survenu entre 1300 et 1900 a fortement enrayé l'autonomie réciproque des provinces historiques, surtout au centre du domaine d'Oïl, tout en repoussant la plupart des différences interponctuelles majeures encore subsistantes aux marges périphériques (Picardie, Wallonie, Lorraine, Franche-Comté). Cette constatation cadre fort bien avec ce que nous savons de l'action pluriséculaire de l'irradiation (circulaire) du type linguistique francien aux dépens des autres types linguistiques d'Oïl.

2.4. *Plateforme comparative 4: la synopse des "coefficients d'asymétrie de Fisher (CAF)"*

Voir les cartes 13-18.

La synopse des coefficients d'asymétrie de Fisher (CAF)[31] constitue – à cause de son excellente utilité en matière de diachronie – une des pièces fortes de la panoplie dialectométrique entière. C'est qu'elle permet de saisir, par voie quantitative, une des propriétés centrales de n'importe quel réseau interactif, à savoir le degré d'interaction avec lequel chaque élément (ou membre) du réseau respectif, participe aux flux d'interactions (ou d'échanges) qui, inévitablement, sillonnent le maillage du réseau en question.

Evidemment, il y aura, dans un réseau donné, toujours des éléments dont la participation aux dynamismes internes sera grande alors que d'autres éléments resteront à l'écart et feront figure – métaphoriquement parlant – d'«abstentionnistes» ou de «trouble-fête». Du point de vue linguistique, les phénomènes d'interaction saisis par le CAF correspondent à ce qu'il est convenu d'appeler, en matière de linguistique, « brassage » (ou aussi: « compromis ») linguistique (ou « Sprachausgleich » en allemand). L'intensité du *brassage linguistique* sera considérable là où la plupart des attributs linguistiques locaux appartiennent à des aires dialectales très grandes ou en voie d'élargissement, alors que l'inverse sera vrai pour les éléments du réseau dont les attributs linguistiques locaux font partie d'aires dialectales plutôt restreintes ou en voie de rétrécissement. Dans cette situation, le défi central que le

[30] Pour des cartes à cloisons relatives au réseau intégral de l'ALF (et visualisées suivant des modalités différentes) voir Goebl 2002, 36 ; 2002, 40 ; 2003, 98-99 et 2004, 280.
[31] Pour la formule du CAF et d'autres détails statistiques y ayant trait, cf. Goebl 1981, 394-401; 1984 I, 150-153 et 2003, 81-84.

©Universitat de Barcelona

dialectométricien doit relever, consiste à mesurer le taux quantitatif de l'enchevêtrement (ou de l'imbrication) d'aires dialectales majeures et mineures.

Les scores du CAF se calculent à l'aide de deux paramètres classiques (*moyenne arithmétique* et *écart-type*) de la distribution de fréquence (ou de similarité) respective et se situent des deux côtés de la valeur zéro. Alors que les valeurs négatives (symbolisées, sur les cartes 13-18, en bleu foncé) correspondent à des zones à forte interaction (ou connectivité) linguistique, les valeurs positives (coloriées en rouge sur les cartes 13-18) renvoient à des régions où l'interaction et le contact linguistique avec le reste du réseau sont relativement faibles.

2.4.1. Interprétation des cartes 13-14

Il s'agit de deux analyses reposant chacune sur le corpus intégral respectif (Dees ou ALF). Sur la carte 13, située à gauche et relative à la fin du 13^{e} siècle, les zones rouges[32] (et, partant, « isolationnistes », c'est-à-dire disposant encore d'une certaine individualité régionale) se situent toutes à la périphérie du domaine d'Oïl: l'on y discerne, en effet, le Poitou, le Maine, la Normandie, le Hainaut et la Wallonie, la Lorraine et la Franche-Comté. Les plages bleues par contre, caractérisées par une connectivité linguistique beaucoup plus intense, occupent les parties centrales du domaine d'Oïl tout en formant des zones-tampon émergeant entre les provinces historiques mentionnées ci-dessus.

Sur la carte 14, relative aux dialectes parlés vers 1900, la zone bleue (ouverte à l'interaction et riche en contacts linguistiques de toute sorte) a fait tache d'huile tout en repoussant à l'extrême périphérie ses contre-parties à faible interaction linguistique. A remarquer, dans le Poitou, le polygone rouge isolé qui, des paysages scripturaires occidentaux du Moyen Age, constitue la dernière survivance.

2.4.2. Interprétation des cartes 15-16 et 17-18

Evidemment, il serait intéressant d'étudier le phénomène d'expansion rendu visible sur les cartes 13 et 14, aussi à l'aide de corpus *partiels* relatifs, par exemple, au *vocalisme* ou au

[32] Pour mieux faire ressortir le message central des profils choroplèthes des cartes 13-18, nous avons choisi une visualisation avec seulement deux paliers chromatiques.

©Universitat de Barcelona

consonantisme respectivement[33]. Ce faisant, l'on pourrait déterminer la différence qui, éventuellement, existerait entre la rapidité avec laquelle l'espace du domaine d'Oïl aurait été homogénéisé, au cours de six siècles, sur le plan di vocalisme d'un côté et sur celui du consonantisme de l'autre. Or, la leçon des cartes 15-16 et 17-18 est très claire à ce propos. Alors que, pour le seul vocalisme (voir les cartes 15 et 16), l'effet égalisateur, jailli au centre du domaine d'Oïl et localisé aux points de jonction entre plusieurs provinces historiques, correspond de très près à ce qui vaut pour le corpus intégral (voir les cartes 13 et 14), tout autre est la situation pour le consonantisme dont la fortune diachronique semble avoir été quelque peu différente.

C'est que les cartes 17 et 18 nous suggèrent que l'aménagement consonantique de l'espace du domaine d'Oïl n'a pratiquement pas changé entre 1300 et 1900 et que, de ce fait, la vitesse de l'évolution diachronique du consonantisme est restée de loin inférieure à celle du vocalisme (voir les cartes 15 et 16).

2.5. *Plateforme comparative 5: la corrélation spatiale entre les aménagements phonétique et consonantique du domaine d'Oïl*

Voir les cartes 19 et 20.

La question de l'évolution diachronique inégale du vocalisme et du consonantisme peut être encore approfondie par l'application de la dialectométrie corrélative. C'est qu'il est possible de calculer, pour les corpus de Dees et de l'ALF et à partir de sous-corpus préalablement définis, deux matrices de similarité, l'une relative au *vocalisme* et l'autre au *consonantisme*, de les corréler par la suite et d'en visualiser les résultats: voir à ce sujet, une fois de plus, la figure 4. La mise en carte de ces calculs[34] est très pertinente; elle confirme pleinement l'interprétation que nous avons donnée, dans le paragraphe précédent, des deux vitesses avec lesquelles le changement linguistique s'est déroulé en matière de vocalisme et de consonantisme.

[33] Pour éliminer les effets perturbateurs qu'auraient eus les données manquantes sur la comparaison des deux sous-corpus en question, nous nous sommes servis, dans le cadre de cette analyse, des sous-corpus réduits (RED) du vocalisme (avec seulement 119 cartes de travail) et du consonantisme (avec seulement 66 cartes de travail). Les mêmes sous-corpus réduits ont été utilisés pour la confection de la carte 19, présentée dans le paragraphe 2.5.

[34] Le choix de quatre paliers chromatiques pour la visualisation des cartes 19 et 20 s'explique à nouveau par la recherche d'une expressivité iconique aussi grande que possible.

©Universitat de Barcelona

La structuration choroplèthe de la carte 19 (relative aux données de l'atlas de Dees) est très claire: les polygones en bleu foncé – équivalents aux zones où le vocalisme et le consonantisme divergent le plus – occupent les endroits ou, au 13e siècle – au moins dans l'optique de l'atlas scripturaire de Dees – , les sous-catégories linguistiques du vocalisme et du consonantisme n'allaient pas encore « de pair ». L'agencement iconique de la carte 20 par contre, dont la structuration est complètement opposée à celle de la carte 19, nous signale que la « mise au pas » du vocalisme et du consonantisme est chose faite dans les centres propulseurs situés au coeur du domaine d'Oïl, et que les ultimes divergences majeures ne subsistent qu'à l'extrême périphérie (Normandie, Hainaut, Wallonie, Franche-Comté). Cette réorientation spatiale (ou « synchronisation ») des deux catégories linguistiques de base, opérée entre 1300 et 1900, ne contredit aucunement les grandes lignes de nos connaissances traditionnelles sur l'histoire de la langue française.

2.6. *Plateforme comparative 6: l'analyse dendrographique*

Voir les cartes 21-24.

Du point de vue historique, l'analyse dendrographique (dite aussi analyse « génétique » ou « généalogique ») a, en matière de linguistique historique, une très longue tradition. Celle-ci débute avec l'indo-européaniste allemand August Schleicher, qui, en 1863, a proposé un schéma dendrographique pour la classification généalogique des langues indo-européennes tout en se servant expressément de visées et méthodes empruntées à la biologie du temps. Depuis lors, de telles expériences ont eté réitérées d'innombrables fois, le plus souvent de la part de comparatistes, lexicostatisticiens et de spécialistes de glottochronologie qui, en l'occurrence, se sont servis très souvent aussi de méthodes quantitatives[35]. L'utilité heuristique d'un schéma dendrographique est double: c'est qu'il concerne tant la diachronie que la synchronie. Il peut donc servir d'un côté à la reconstruction (théorique) d'étapes diachroniques antérieures à celle du relèvement des données analysées, et fournir de l'autre une classification hiérarchique en synchronie, qu'elle soit spatiale (comme c'est le cas ici) ou non.

Les arborescences générées par la taxométrie (le plus souvent d'inspiration biologique) ont les propriétés formelles suivantes:

[35] Voir à ce sujet l'excellent volume-recueil de Hoenigswald/Wiener (1987).

©Universitat de Barcelona

a) Il s'agit d'une hiérarchie exclusivement binaire de classes (clusters, groupes etc.) disjointes et non recouvrantes.

b) Les clusters du graphe disposent d'une hétérogénéité quantitative interne différente suivant leur position à l'intérieur de l'arbre: plus tel cluster se rapproche de la racine (ou du tronc) de l'arbre, plus son hétérogénéité quantitative interne augmente.

La taxométrie moderne dispose, depuis quelque 50 à 60 ans, d'une riche panoplie d'algorithmes classificatoires parmi lesquels nous avons choisi l'algorithme proposé par le statisticien américain Joe Ward, Jr. (voir les cartes 21 et 22) et l'algorithme dit « Complete Linkage » (voir les cartes 23 et 24) tout en les appliquant aux données de l'atlas de Dees et de l'ALF[36].

Les deux algorithmes servent à l'établissement d'une « classification ascendante hiérarchique » (CAH). En recourant aux données numériques stockées dans la matrice de similarité, les algorithmes-CAH commencent par fusionner, parmi les N éléments sujets à classification, ceux dont la similarité quantitative (i. e. dialectométrique) est la plus grande tout en engendrant – au cours de l'avancement, dans la direction des feuilles vers la racine de l'arbre, de ces fusions effectuées toujours par paires – un graphe hiérarchique binaire disposant de N « feuilles », de N-1 bifurcations (ou ramifications, embranchements) et d'une racine (ou d'un tronc) unique.

Pour l'interprétation dialectométrique des arbres calculés (et dûment visualisés par la suite), il faut en transposer le rendement classificatoire dans l'espace respectif. Ce processus est appelé « spatialisation ». En regardant les cartes 21 et 22, l'on constate qu'à la première ramification des deux arbres fait écho, dans l'espace, une bipartition du domaine d'Oïl dans la direction est-ouest, et aussi que les embranchements successifs créent des classes (ou « dendrèmes »)[37] dans le « feuillage » de l'arbre dont les équivalents spatialisés (« chorèmes ») constituent toujours des aires disposant d'une cohérence diatopique surprenante.

Il va de soi que ni pour les deux algorithmes utilisés ni pour les deux stocks de données examinés, l'on n'est en droit de s'attendre à des résultats classificatoires parfaitement

[36] Pour l'algorithme de Ward cf. Sneath/Sokal 1973, 204 s., Bock 1974, 407 s. et Chandon/Pinson 1981, 123-124; pour l'algorithme dit « Complete Linkage » cf. Bock 1974, 392 s., Sneath/Sokal 1973, 222 s. et Chandon/Pinson 1981,106s.

[37] Les termes « dendrèmes » et « chorèmes » désignent des groupements jugés importants (ou « significatifs ») situés, respectivement, à l'intérieur de l'*arbre* ou de sa *projection* dans l'*espace* (spatialisation). Pour l'historique de cette terminologie combinée cf. Goebl 2003, 85 (note 13).

©Universitat de Barcelona

identiques. L'on constate pourtant que les classifications obtenues obéissent, en ligne générale, toujours aux mêmes tendances (hiérarchiques et spatiales), ce qui prouve, une fois de plus, que les différentes structures de profondeur identifiables dans l'atlas de Dees et dans l'ALF – outre d'être solidement ancrées dans les données respectives – se ressemblent de très près. Il est également intéressant de voir que l'individualité diatopique de certains paysages linguistiques traditionnels, telles la Picardie, la Wallonie, la Lorraine et les provinces du Sud-Ouest, ressort fort bien aussi par le biais d'une CAH.

3. Conclusion

Des analyses dialectométriques que nous venons de présenter, il appert avec netteté que les deux stocks de données examinés (l'atlas de Dees et l'ALF) sont riches en structurations diatopiques de toute sorte et que celles-ci se manifestent soit « en superficie » soit « en profondeur ». Jetons, pour illustrer encore une fois la différence entre les termes métaphoriques de « superficie » et de « profondeur », un dernier coup d'oeil sur les dimensions des matrices de similarité utilisées et sur l'importance variable de leur apport aux diverses analyses dialectométriques présentées. Selon la formule combinatoire N/2 (N-1)[38], nos matrices de similarité disposent – pour N = 85 – de 3570 scores de similarité différents. Ceci signifie que, pour une seule carte de similarité basée sur 85 scores numériques, le taux d'exhaustion est de 2,85% (= 85 : 3570) alors que le même taux d'exhaustion s'élève à 6,30% (= 225 : 3570) pour une carte à cloisons qui repose sur 225 valeurs particulières. La même analyse nous dit cependant que le calcul du coefficent d'asymétrie de Fisher (CAF), des cartes à corrélations ou des arbres généalogiques ne peut se faire qu'à l'aide du total des 3570 scores disponibles. Il en résulte qu'il existe, parmi les différentes espèces de l'analyse dialectométrique, une certaine hiérarchie taxométrique à laquelle nous avons appliqué les concepts métaphoriques de « superficie » et de « profondeur ».

Cet état de choses est d'ailleurs bien loin d'être l'apanage exlusif de la dialectométrie. Il est courant également dans toutes les disciplines quantifiantes où, pour l'analyse exploratoire des données empiriques et pour l'acceptation ou le rejet de certaines hypothèses ou de

[38] Pour une matrice de données avec N éléments (et p attributs), une comparaison de ses éléments par paires aboutit à une matrice de similarité carrée avec les dimensions N^2. Comme la diagonale de cette matrice carrée se compose toujours de scores avec la valeur 1 (ou 100%) et que les deux moitiés de la matrice carrée figurant des deux côtés de la diagonale ne contiennent que des scores parfaitement identiques, le nombre de scores utilisables à des fins classificatoires se calcule comme suit: $(N^2 – N): 2 = N/2 (N – 1)$.

©Universitat de Barcelona

théories de recherche, l'on fait recours à des procédés mathématiques ou statistiques. Toujours est-il qu'à la lumière de telles analyses, notre conjecture de départ selon laquelle l'aménagement linguistique du domaine d'Oïl des 13^{e} et 19^{e} siècles repose sur des prérequis très similaires, a pu être confirmée à plusieurs reprises. Il serait hautement souhaitable de réitérer cette expérience, soit pour le domaine d'Oïl et à l'aide de données nouvelles, soit pour un autre domaine de la Romania et à l'aide de données déjà disponibles.

4. Remerciements

La rédaction de cette contribution ainsi que tous les travaux préparatoires y ayant trait présupposent la coopération concertée d'une équipe bien rodée. C'est pourquoi je tiens à remercier très cordialement tous ceux qui, de près ou de loin, ont apporté leur contribution à la recherche que nous venons de présenter.

4.1. Confection des figures et des cartes de cet article, entretien permanent du logiciel et du matériel de VDM:
Slawomir Sobota (Salzbourg),

4.2. Création et mise à jour périodique du logiciel VDM[39]:
Edgar Haimerl (Blaustein, Allemagne),

4.3. Taxation des données de l'ALF:
Barbara Aigner, Irmgard Dautermann, Hildegund Eder, Susanne Oleinek, Annette Ida Schatzmann (Salzbourg),

4.4. Toilette stylistique de mon texte français:
Lily Ditz-Fuhrich (Salzbourg).

4.5. Appui financier de nos recherches dialectométriques: projets 12 414 et 13 349 du Fonds national autrichien pour la recherche scientifique FWF (« Fonds zur Förderung der wissenschaftlichen Forschung in Österreich »).

[39] Consu Consulter aussi notre site internet bilingue (allemand-anglais) relatif à la dialectométrie et au logiciel VDM: http://ald.sbg.ac.at/dm/. Voir aussi le site suivant : http://www.dialectometry.com/.

©Universitat de Barcelona

5. Références bibliographiques et abréviations techniques souvent utilisées

5.1. *Abréviations techniques souvent utilisées*

CAF	Coefficient d'asymétrie de Fisher	Cf. le paragraphe 2.4.
CAH	classification ascendante hiérarchique	Cf. le paragraphe 2.6.
CONS	Corpus consonantique	Cf. les paragraphes 1.1 et 1.2.
DEM_{jk}	Distance Euclidienne Moyenne	Voir le paragraphe 1.4. et la figure 3.
$IPI(1)_{ik}$	Indice d'Identité Pondéré (avec le poids 1)	Cf. le paragraphe 1.4.
IRI_{jk}:	Indice Relatif d'Identité	Cf. le paragraphe 1.4.
$r(BP)_{ik}$	coefficient de corrélation de Bravais et Pearson	Cf. les paragraphes 2.1. et 2.2.
RED	Corpus réduit	Cf. les paragraphes 1.1 et 1.2.
SEM_{jk}	Similarité Euclidienne Moyenne	Voir le paragraphe 1.4. et la figure 3.
TOT	Corpus intégral	Cf. les paragraphes 1.1 et 1.2.
VDM	Visual DialectoMetry, logiciel dialectométrique créé par Edgar Haimerl	Cf. le paragraphe 1.
VOC	Corpus vocalique	Cf. les paragraphes 1.1 et 1.2.

5.2. *Références bibliographiques*

ALF: Gilliéron, Jules/Edmont, Edmond (eds.) (1902-1910) *Atlas linguistique de la France*, Paris: Champion, 10 vol. (réimpression: Bologne: Forni, 1968).

BESCH, Werner (1985) "Die Entstehung und Ausformung der neuhochdeutschen Schriftsprache/Standardsprache", in Besch, Werner/Reichmann, Oskar/Sonderegger, Stefan (éds.): *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*, Berlin, New York: Walter de Gruyter, vol. I, 1781-1810.

BOCK, Hans Hermann (1974) Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse), Göttingen: Vandenhoeck & Ruprecht.

CHANDON, Jean-Louis/Pinson, Suzanne (1981) *Analyse typologique. Théories et applications*, Paris, New York, Barcelone, Milan: Masson.

DEES, Anthonij (1980) *Atlas des formes et des constructions des chartes françaises du 13^{e} siècle*, Tübingen: Niemeyer (Beihefte zur Zeitschrift für romanische Philologie, vol. 178).

©Universitat de Barcelona

DEES, Anthonij (1984) "Regards quantitatifs sur les variations régionales en ancien français", in: Goebl, Hans (éd.): *Dialectology*, Bochum: Brockmeyer (Quantitative Linguistics, vol. 21), 102-120.

DEES, Anthonij (1987) *Atlas des formes linguistiques des textes littéraires de l'ancien français*, Tübingen: Niemeyer (Beihefte zur Zeitschrift für romanische Philologie, vol. 212).

GOEBL, Hans (1981) "Eléments d'analyse dialectométrique (avec application à l'AIS)", in: *Revue de linguistique romane,* 45, 349-420.

GOEBL, Hans (1982) "Compte rendu de Dees 1980", in *Zeitschrift für französische Sprache und Literatur,* 92, 280-283.

GOEBL, Hans (1983) "Parquet polygonal et treillis triangulaire: les deux versants de la dialectométrie interponctuelle", in *Revue de linguistique romane*, 47, 353-412.

GOEBL, Hans (1984) *Dialektometrische Studien. Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, Tübingen: Niemeyer (Beihefte zur Zeitschrift für romanische Philologie, vol. 191-193).

GOEBL, Hans (1987) "Points chauds de l'analyse dialectométrique: pondération et visualisation", in *Revue de linguistique romane,* 51, 63-118.

GOEBL, Hans (1989) "-ORE(A) statt -ORIU(A) in der Normandie: einem Suffixwechsel mit skriptologischen Mitteln auf der Spur", in Ernst, Gerhard/Stefenelli, Arnulf (éds.): *Studien zur romanischen Wortgeschichte. Festschrift für Heinrich Kuen zum,* 90. Geburtstag, Stuttgart: Steiner, 92-102.

GOEBL, Hans (1993) "Compte rendu de Dees 1987", in *Zeitschrift für französische Sprache und Literatur,* 103, 185-187.

GOEBL, Hans (1998) "Zu einer dialektometrischen Analyse der Daten des Dees-Atlasses von 1980", in: Werner Edeltraud/Liver, Ricarda/Storck, Yvonne/ Nicklaus, Martina (éds.): *Et multum et multa. Festschrift für Peter Wunderli zum 60*. Geburtstag, Tübingen: Narr, 293-309.

GOEBL, Hans (2000) "La dialectométrisation de l'ALF: présentation des premiers resultants", in: *Linguistica*, 40, 209-236.

GOEBL, Hans (2001) "La strutturazione geolinguistica del dominio d'oïl nel 13° secolo e alla fine dell'Ottocento. Un raffronto dialettometrico", in: Zamboni, Alberto/Del Puente, Patrizia/Vigolo, Maria Teresa (éds.): *La dialettologia oggi fra tradizione e nuove*

©Universitat de Barcelona

metodologie. Atti del Convegno internazionale, Pisa 10-12 Febbraio 2000, Pisa: Edizioni ETS, 11-43.

GOEBL, Hans (2002) "Analyse dialectométrique des structures de profondeur de l'ALF", in: *Revue de linguistique romane*, 66, 5-63 (avec 24 cartes à couleurs).

GOEBL, Hans (2003) "Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur", in: *Estudis Romànics,* 25, 59-96 (avec 24 cartes à couleurs).

GOEBL, Hans (2004) "Sprache, Sprecher und Raum: eine kurze Darstellung der Dialektometrie. Das Fallbeispiel Frankreich", in: *Mitteilungen der österreichischen geographischen Gesellschaft,* 146, 247-286.

GOEBL, Hans & Guillaume SCHILTZ (2001) "Der "Atlas des formes et des constructions des chartes françaises du 13^{e} siècle" von Anthonij Dees (1980) – dialektometrisch betrachtet", in: Gärtner, Kurt/Holtus Günter/Rapp, Andrea/Völker, Harald (éds.): *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.

HOENIGSWALD, Henry M. & Linda F WIENER (éds.) (1987) *Biological Metaphor and Cladistic Classification. An Interdisciplinary Perspective,* Philadelphia: University of Pennsylvania Press.

HUBER, Onno & Karin VAN REENEN-STEIN (1988) "Corrélations et groupements dans l'Atlas des formes et des constructions des chartes françaises du XIII^{e} siècle", in: van Reenen, Pieter/van Reenen-Stein, Karin (éds.): *Distributions spatiales et temporelles, constellations de manuscrits. Etudes de variation linguistique offertes à Anthonij Dees à l'occasion de son 60^{me} anniversaire*, Amsterdam, Philadelphia: Benjamins, 93-101.

REMACLE, Louis (1992) La différenciation dialectale de la Belgique romane avant 1600, Genève: Droz.

SCHILTZ, Guillaume (1997) *Die Anwendung künstlicher neuronaler Netze zur Klassifikation altfranzösischer Dialekte* [Amsterdam, Second International Congress of Dialectologists and Geolinguists, 1997], téléchargeable sous l'adresse suivante: http://pages.unibas.ch/anglist/schiltz/statistics.htm.

SCHLEICHER, August (1863) Die Darwinsche Theorie und die Sprachwissenschaft. Offenes Sendschreiben an Herrn Dr. Ernst Häckel, a. o. Professor der Zoologie und Director des zoologischen Museums an der Universität Jena, Weimar: Böhlau.

©Universitat de Barcelona

SÉGUY, Jean (1971) La relation entre la distance spatiale et la distance lexicale, in: Revue de linguistique romane, 35, 335-357.

SÉGUY, Jean (1973) "La fonction minimale du dialecto", in: Straka, Georges/Gardette, Pierre (éds.): *Les dialectes romans de France à la lumière des atlas régionaux* (Colloque de Strasbourg, 1971), Paris: Editions du CNRS, 27-37.

SNEATH, Peter A., Robert R. Sokal (1973) Numerical Taxonomy. The Principles and Practice of Numerical Classification, San Francisco: Freeman.

VÖLKER, Harald (2003) "Skripta und Variation. Untersuchungen zur Negation und zur Substantivflexion in altfranzösischen Urkunden der Grafschaft Luxemburg (1237-1281)", Tübingen: Niemeyer (Beihefte zur Zeitschrift für romanische Philologie, vol. 305).

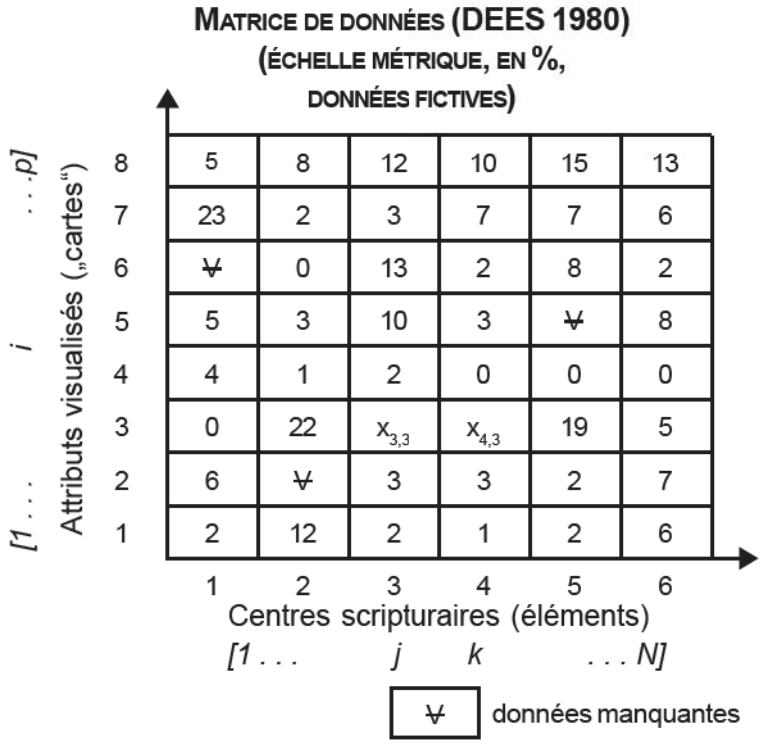
©Universitat de Barcelona

LISTE DES POINTS DE L'ATLAS DE DEES (1980)

MACRO (MA)	NOM	*MICRO (MI)*
1	Charente, Charente-Maritime	1
	La Rochelle et environs	2
2	Vendée, Deux-Sèvres	3
	Vendée	4
	Deux-Sèvres	5
3	Vienne	6
	Mirebeau et environs (= Poitiers)	7
	Loudun et environs	8
	Châtellerault et environs	9
4	Indre, Cher	10
5	Orléanais	11
	Eure-et-Loir	12
	Loir-et-Cher	13
6	Indre-et-Loire	14
	Tours et environs	15
	Loches et environs	16
7	Maine-et-Loire	17
	Maine-et-Loire sud + Angers	18
8	Mayenne, Sarthe	19
	Le Mans et environs	20
9	Bretagne	21
10	Normandie	22
	Manche	23
	Eure	24
	Seine-Maritime	25
11	Somme, Pas-de-Calais	26
	Somme centre + sud	27
	Somme nord-ouest	28
	Pas-de-Calais sud-est	29
	Pas-de-Calais sud-ouest	30
	Pas-de-Calais centre + sud	31
12	Oise	32
	Oise sud-ouest	33
	Oise nord-ouest	34
	Oise nord-est	35
	Oise sud-est	36
13	Aisne	37
	Saint-Quentin et environs	38
14	Nord	39
	Lille et environs	40
	Douai et environs	41

MACRO (MA)	NOM	*MICRO (MI)*
15	Hainaut	42
	Tournai et environs	44
16	Wallonie	45
	Liège (province)	46
	Luxembourg	47
17	Ardennes	48
	Ardennes nord	49
	Ardennes sud	50
18	Marne	51
	Marne ouest	52
	Marne est	53
19	Région parisienne	54
	Val d'Oise	55
	Paris	56
	Seine-et-Marne	57
20	Yonne	58
21	Aube	59
22	Haute-Marne	60
	Langres et environs	61
	Châteauvillain et environs	62
	Chaumont et environs	63
	Haute-Marne nord	64
	Haute-Marne nord-est	65
23	Meuse	66
	Verdun et environs	67
	Meuse sud-ouest	68
	Meuse sud-est	69
24	Moselle, Meurthe-et-Moselle nord	70
	Nancy et environs	71
	Moselle sans Metz	72
	Meurthe-et-Moselle nord	73
	Metz et environs	74
25	Vosges	75
	Neufchâteau	76
26	Franche-Comté	77
	Doubs	78
	Haute-Saône	79
	Jura	80
27	Bourgogne	81
	Dijon et environs	82
	Côte-d'Or nord	83
	Côte-d'Or sud + Saône-et-Loire	84
28	Nièvre, Allier	85

Figure 1 : Liste des unités spatiales (centres scripturaires, points) majeures (Macro) et mineures (Micro) de l'atlas de DEES (1980).



Commentaire: Pour chacun des 298 attributs de son atlas de 1980, Anthonij DEES a défini deux ensembles de graphies complémentaires dont un seul a été mis en carte. En voici un exemple relatif à la carte (attribut graphique) 1 de son atlas: il s'agit des différentes réalisations de l'étymon latin ĔGO:

graphies visualisées: *g*, *ge*, *gie*, etc.
graphies complémentaires non visualisées: *i*, *ie*, *iou*, *je*, *jo*, etc.

Pour le calcul de la fréquence (relative) des graphies visualisées à tel endroit du réseau examiné [%], A. DEES a établi le quotient entre le nombre des chartes disposant l'attribut visualisé (1) et la somme des chartes disposant l'attribut visualisé (1) d'un côté et l'attribut complémentaire non visualisé (2) de l'autre.

Quant à l'exemple cité ci-dessus, les scores respectifs retenus dans la matrice de données se calculent comme suit:

$$\% = 100 * \frac{\sum \text{chartes (du centre scripturaire } j\text{) avec des occurrences de l'attribut visualisé (1)}}{\sum \text{chartes (du centre scripturaire } j\text{) avec des occurrences de l'attribut visualisé (1) et de l'attribut non visualisé (2)}}$$

Les chartes du centre scripturaire j qui ne disposent d'aucun succédané de l'étymon latin ĔGO, restent exclues du calcul.

Figure 2: Tableau explicatif de l'établissement de la matrice de données de l'atlas de Dees (1980). Voir aussi la figure 3.

©Universitat de Barcelona

CALCUL DE LA SEM_{JK} (SIMILARITÉ EUCLIDIENNE MOYENNE)[1)]

$$SEM_{jk} = 100 - 100 * \sqrt{\frac{1}{\tilde{p}} \sum_{i=1}^{\tilde{p}} (x_{ji} - x_{ki})^2}$$

Explication des symboles:

SEM_{jk}	Similarité Euclidienne Moyenne entre les centres scripturaires *j* et *k*
i	index pour un attribut graphique
j,*k*	index pour deux centres scripturaires comparés entre eux
x_{ji}	pourcentage répertorié dans la matrice de données à l'endroit du centre scripturaire *j* et de l'attribut *i*
$\tilde{p}$	nombre des attributs graphiques disponibles (co-présents) dans les vecteurs *j* et *k* de la matrice de données (à l'exclusion de données manquantes)

[1)] Par rapport à la SEM_{jk}, la „Distance Euclidienne Moyenne" (DEM_{jk}) se calcule comme suit: $SEM_{jk} + DEM_{jk} = 100$.

Vérification, à l'aide de la matrice de données de la figure 2 du calcul de la $SEM_{1,2}$ (entre les centres scripturaires 1 et 2)

- Calcul du terme $\sum_{i=1}^{6} (x_{1,i} - x_{2,i})^2$:

  $(2 - 12)^2 + (0 - 22)^2 + (4 - 1)^2 + (5 - 3)^2 + (23 - 2)^2 + (5 - 8)^2 = 1047$

- A remarquer les données manquantes dans *i* = 2 et 6, d'où $p = \tilde{6}$.

- $SEM_{1,2} = 100 - 100 \sqrt{\frac{1047}{6}}$

- $SEM_{1,2} = 100 - 1320{,}98 = -1220{,}98$

Figure 3: Calcul de la similarité quantitative entre deux centres scripturaires *j* et *k* à l'aide de l'indice SEM_{jk} (Similarité Euclidienne Moyenne). Voir aussi la figure 2.

©Universitat de Barcelona

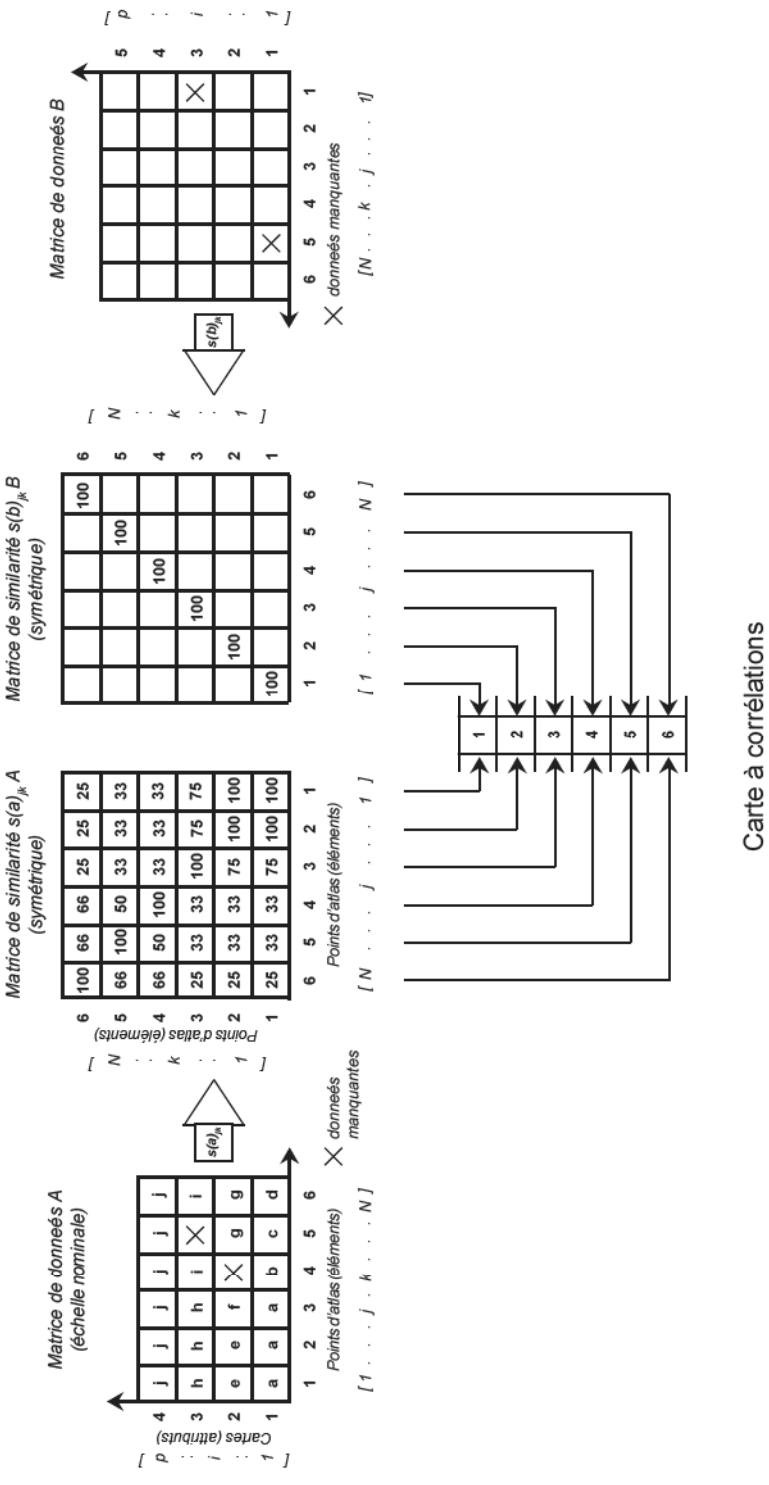
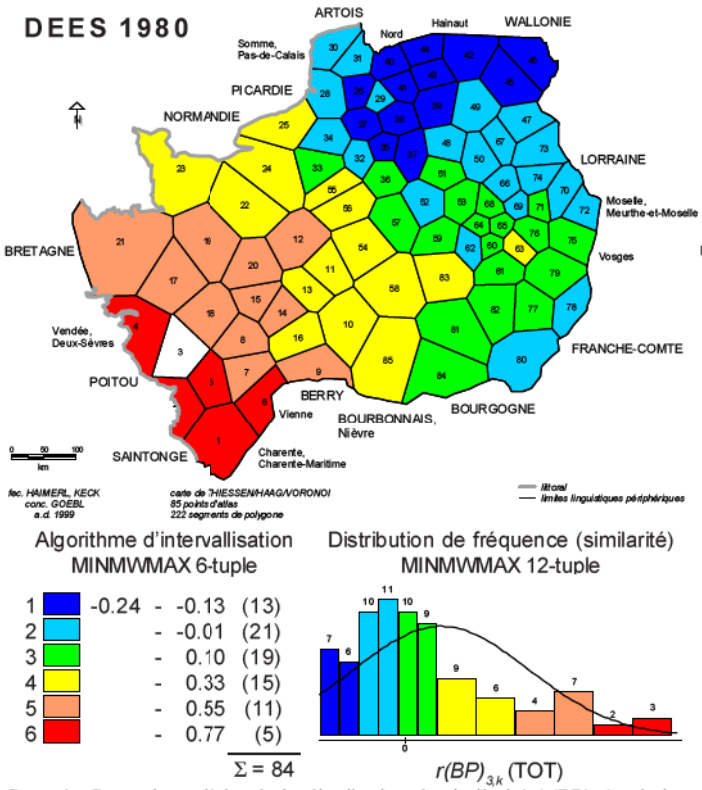
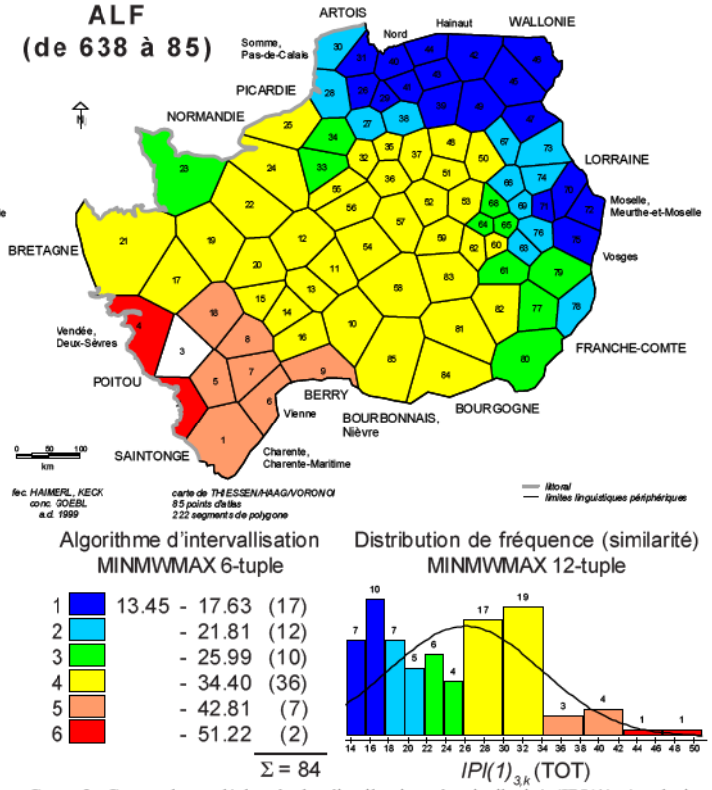


Figure 4: Tableau explicatif de l'établissement de deux matrices de similarité, de leur corrélation successive (à l'aide du coefficient de corrélation de Bravais et Pearson [$r(BP)_{jk}$]) et de l'établissement de la carte à corrélations.

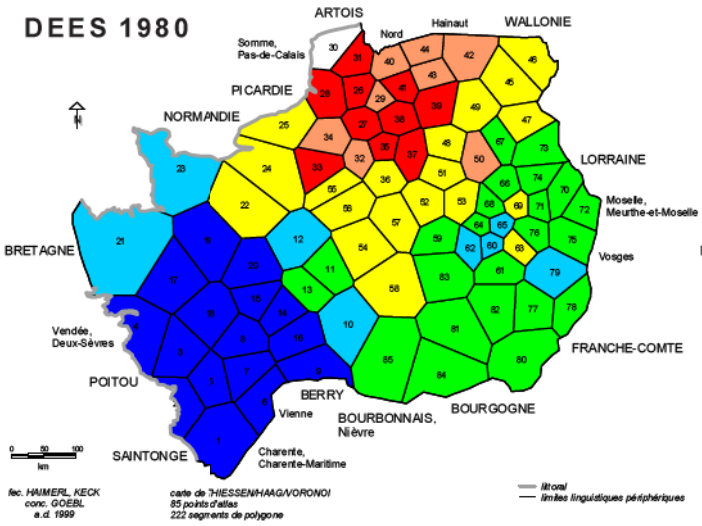
©Universitat de Barcelona

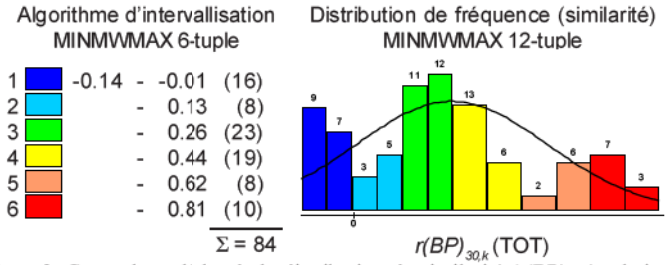


Carte 1: Carte choroplèthe de la distribution de similarité ($r(BP)_{3,k}$) relative au point de référence 3 (DEES 3: Vendée, Deux-Sèvres).
Corpus: Dees 1980, 268 cartes (données intégrales).
Voir aussi le paragraphe 2.1.

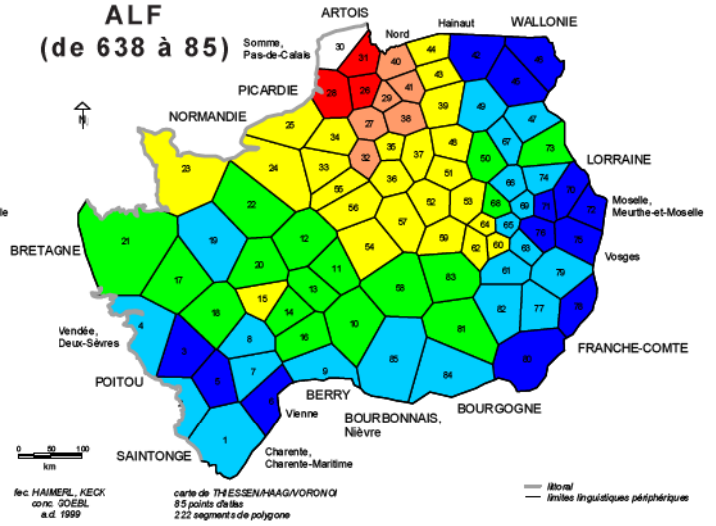


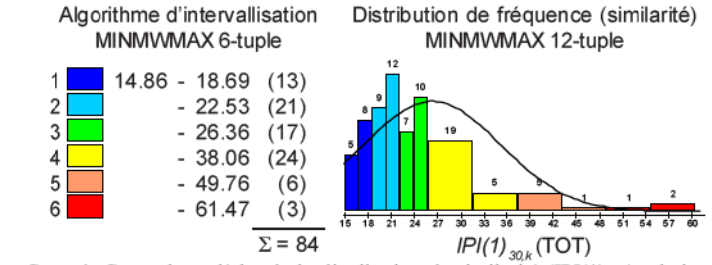
Carte 2: Carte choroplèthe de la distribution de similarité ($IPI(1)_{3,k}$) relative au point de référence 3 (ALF 429: Saint-Germain-le-Prinçay, Vendée).
Corpus: ALF, 1687 cartes de travail (données intégrales).
Voir aussi le paragraphe 2.1.



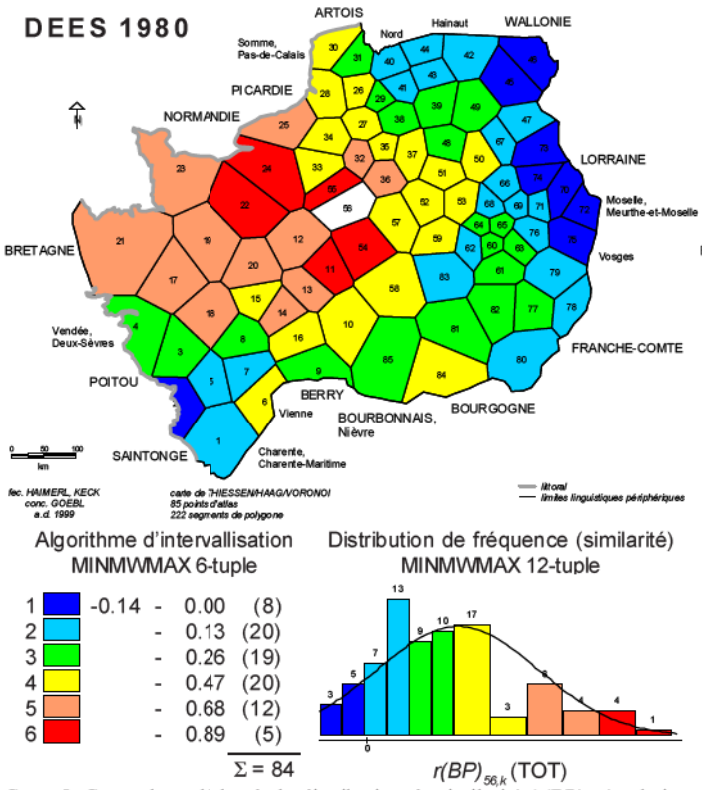


Carte 3: Carte choroplèthe de la distribution de similarité ($r(BP)_{30,k}$) relative au point de référence 30 (DEES 30: Pas-de-Calais, sud-ouest).
Corpus: Dees 1980, 268 cartes (données intégrales).
Voir aussi le paragraphe 2.1.

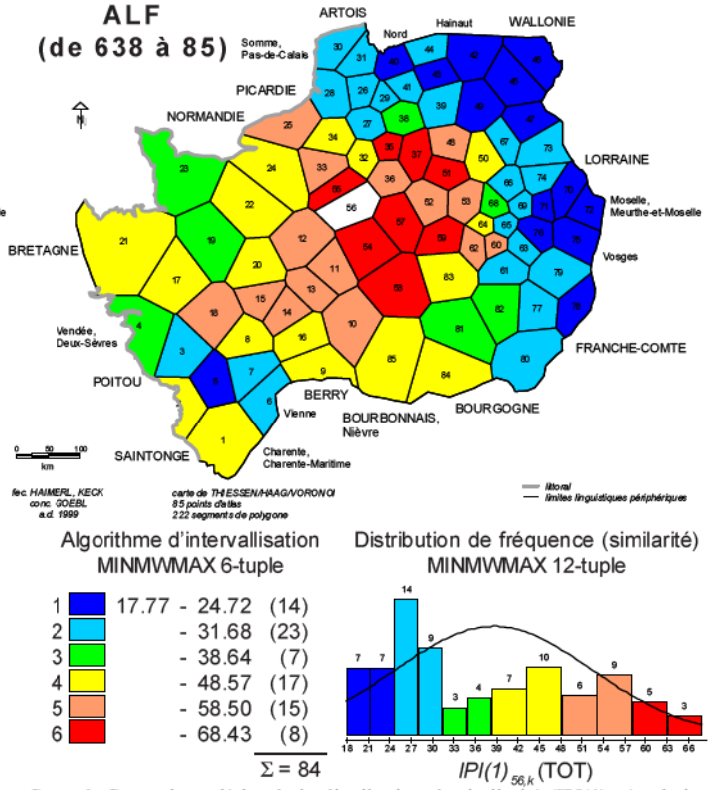




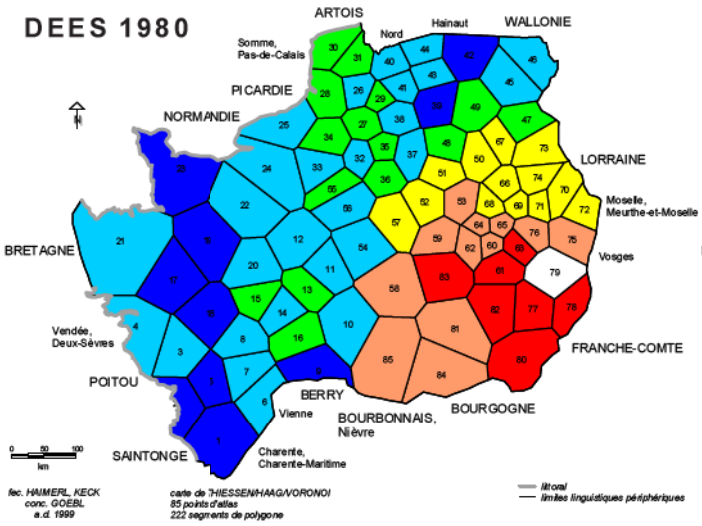
Carte 4: Carte choroplèthe de la distribution de similarité ($IPI(1)_{30,k}$) relative au point de référence 30 (ALF 298: Nort-Leulinghem, Pas-de-Calais).
Corpus: ALF, 1687 cartes de travail (données intégrales).
Voir aussi le paragraphe 2.1.

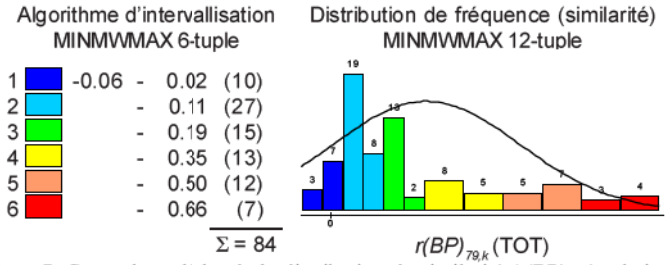


Carte 5: Carte choroplèthe de la distribution de similarité ($r(BP)_{56,k}$) relative au point de référence 56 (DEES 56: Paris).
Corpus: Dees 1980, 268 cartes (données intégrales).
Voir aussi le paragraphe 2.1.

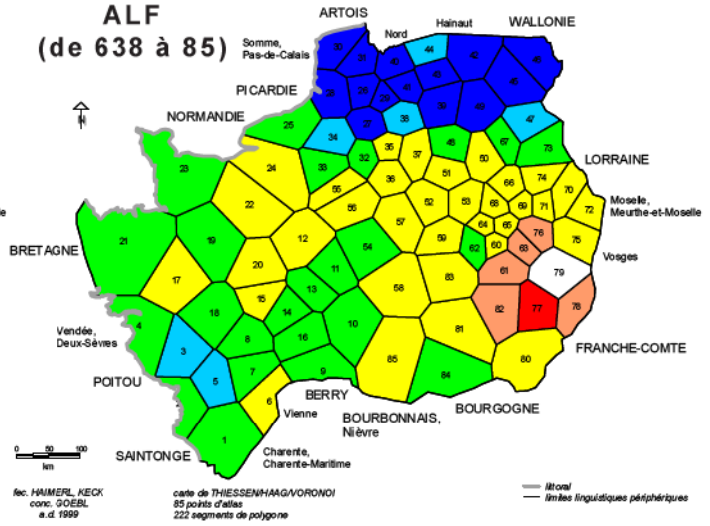


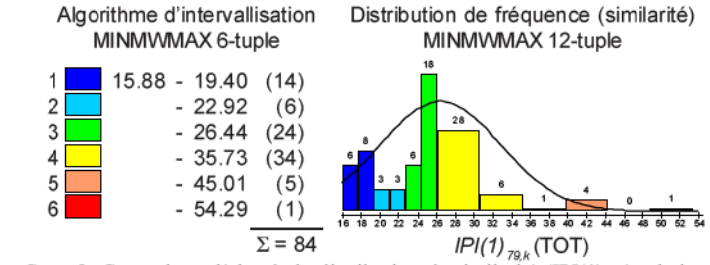
Carte 6: Carte choroplèthe de la distribution de similarité ($IPI(1)_{56,k}$) relative au point de référence 56 (ALF 226: Le Plessis-Piquet; Hauts-de-Seine).
Corpus: ALF, 1687 cartes de travail (données intégrales).
Voir aussi le paragraphe 2.1.



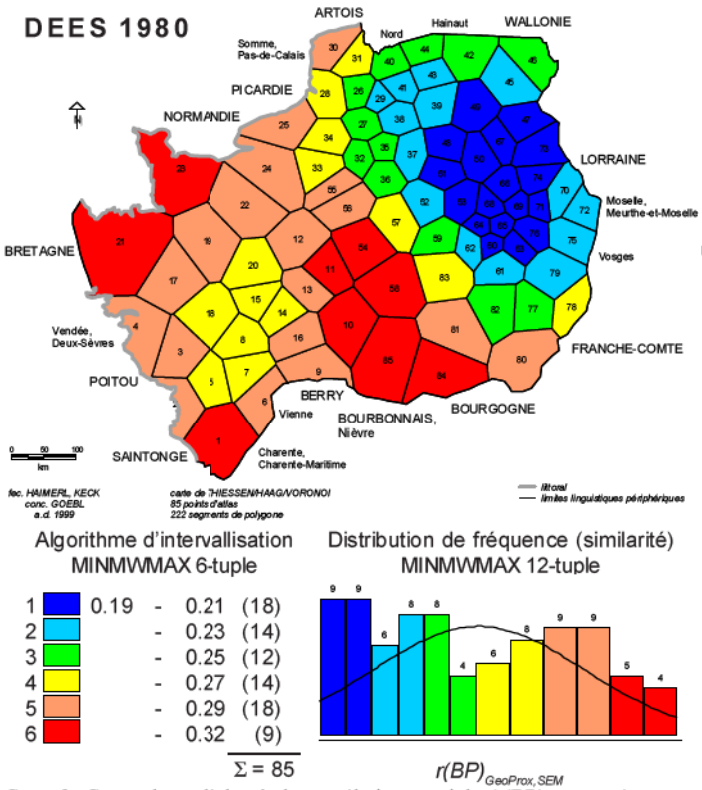


Carte 7: Carte choroplèthe de la distribution de similarité ($r(BP)_{79,k}$) relative au point de référence 79 (DEES 79: Haute-Saône).
Corpus: Dees 1980, 268 cartes (données intégrales).
Voir aussi le paragraphe 2.1.

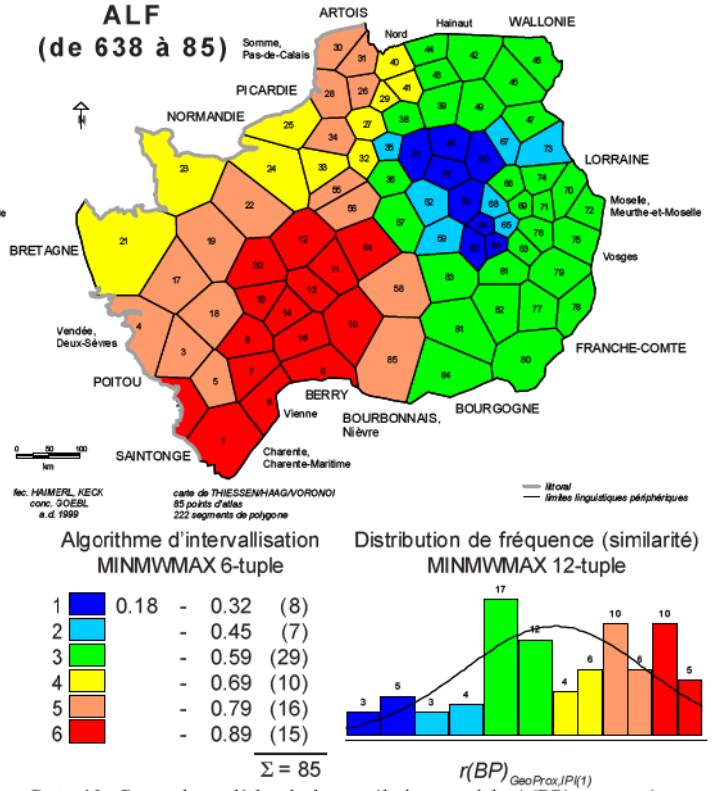




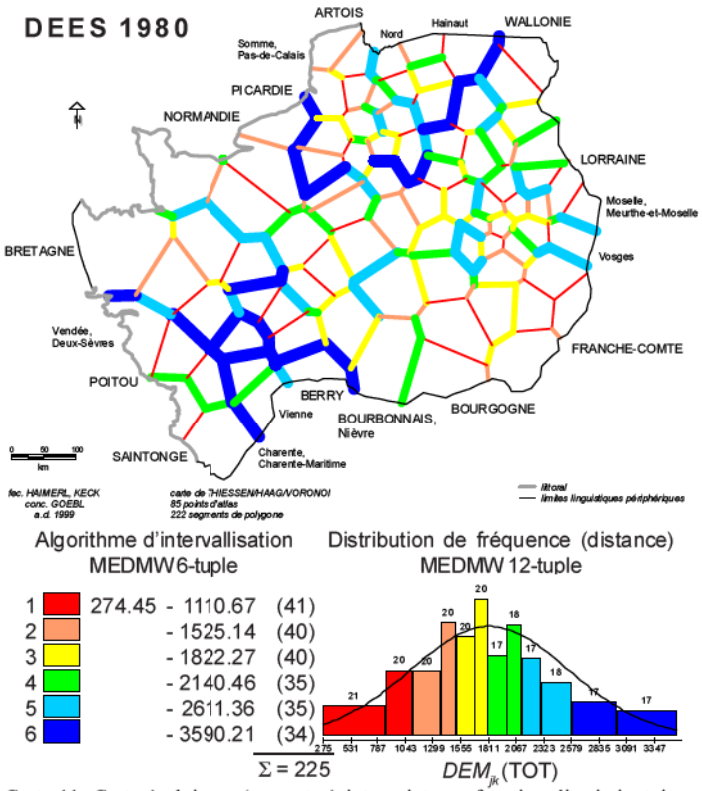
Carte 8: Carte choroplèthe de la distribution de similarité ($IPI(1)_{79,k}$) relative au point de référence 79 (ALF 46: Conflans-sur-Lanterne, Haute-Saône).
Corpus: ALF, 1687 cartes de travail (données intégrales).
Voir aussi le paragraphe 2.1.



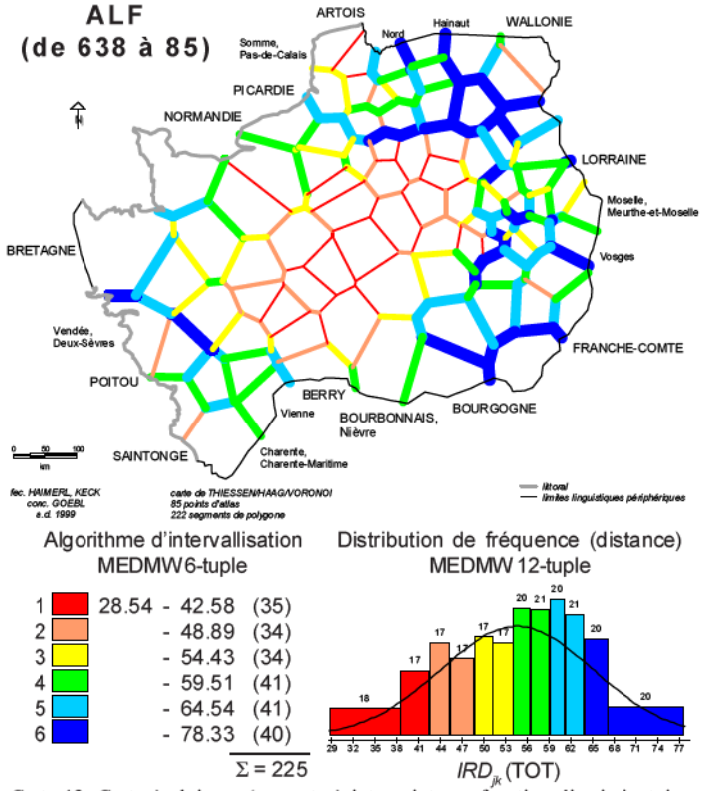
Carte 9: Carte choroplèthe de la corrélation spatiale ($r(BP)_{GeoProx,SEM}$) entre les aménagements euclidien et scripturaire du domaine d'Oïl vers 1300. Voir aussi le paragraphe 2.2.



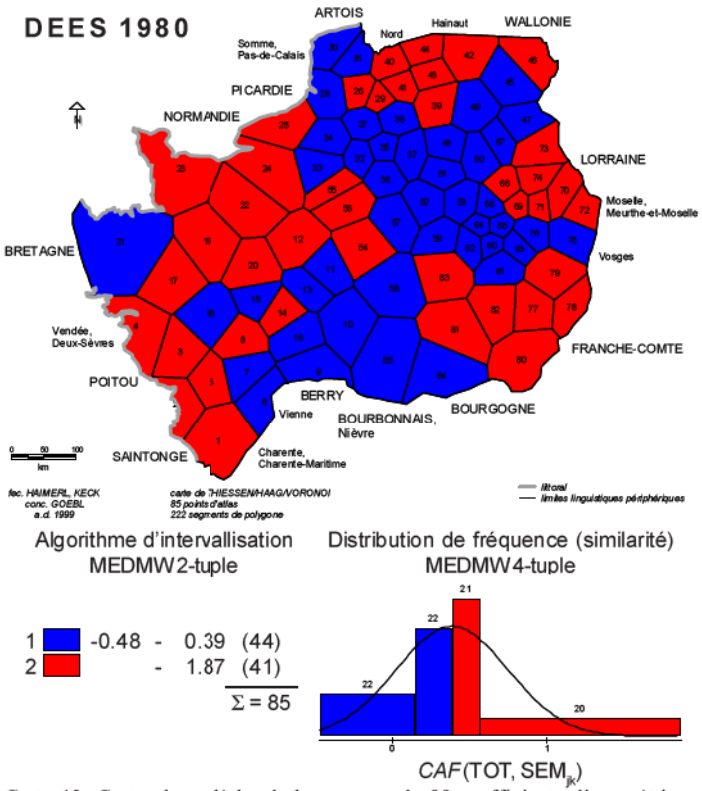
Carte 10: Carte choroplèthe de la corrélation spatiale ($r(BP)_{GeoProx,IPI(1)}$) entre les aménagements euclidien et dialectal du domaine d'Oïl vers 1900. Voir aussi le paragraphe 2.2.



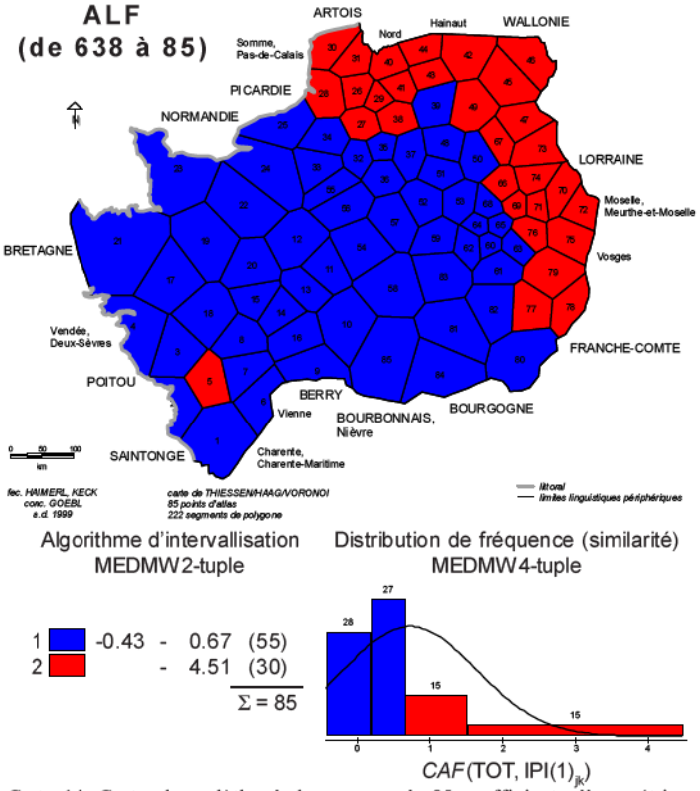
Carte 11: Carte à cloisons (ou carte à interpoints en fonction discriminatoire: isoglosses quantitatives).
Corpus: Dees 1980, 268 cartes (données intégrales).
Indice de distance: DEM_{jk}. Voir aussi le paragraphe 2.3.



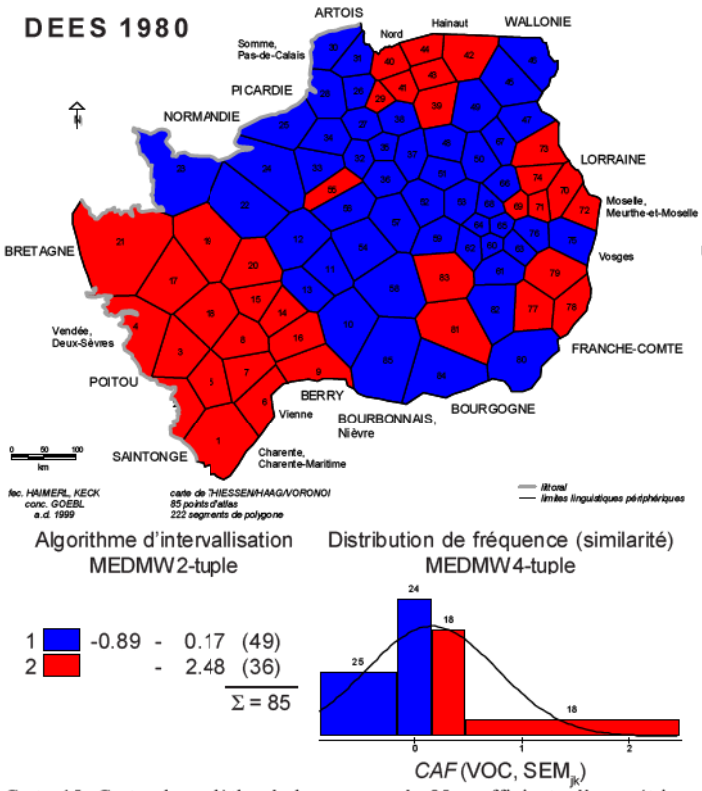
Carte 12: Carte à cloisons (ou carte à interpoints en fonction discriminatoire: isoglosses quantitatives).
Corpus: ALF, 1687 cartes (données intégrales).
Indice de distance: IRD_{jk}. Voir aussi le paragraphe 2.3.



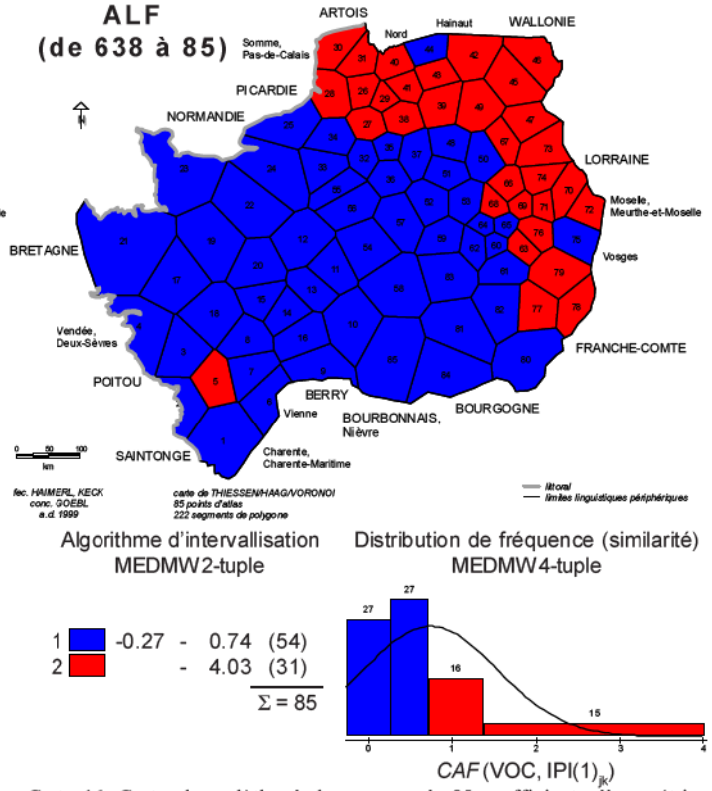
Carte 13: Carte choroplèthe de la synopse de 85 coefficients d'asymétrie de Fisher (CAF).
Corpus: Dees 1980, 268 cartes (données intégrales).
Indice de similarité: SEM_{jk}. Voir aussi le paragraphe 2.4.



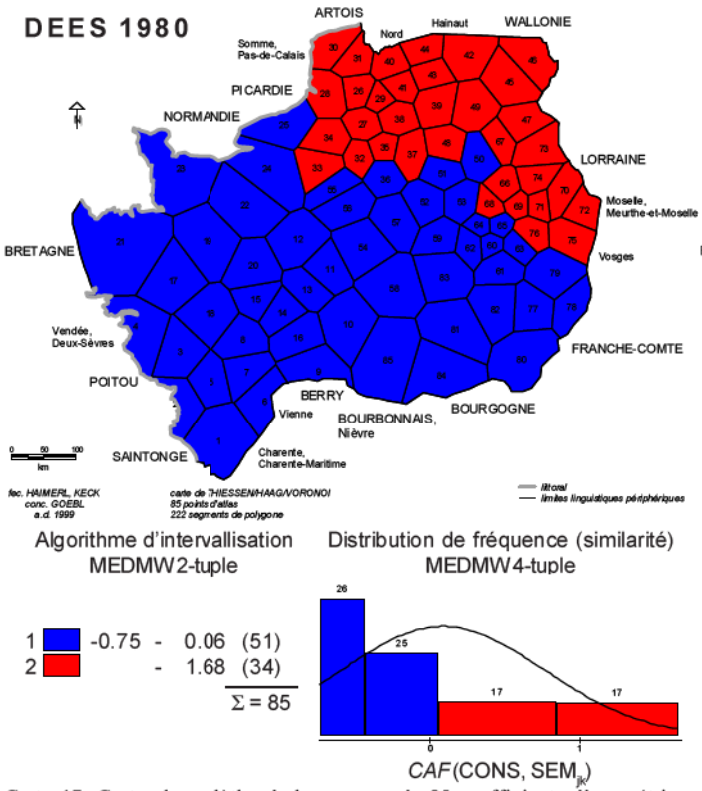
Carte 14: Carte choroplèthe de la synopse de 85 coefficients d'asymétrie de Fisher (CAF).
Corpus: ALF, 1687 cartes de travail (données intégrales).
Indice de similarité: $IPI(1)_{jk}$. Voir aussi le paragraphe 2.4.



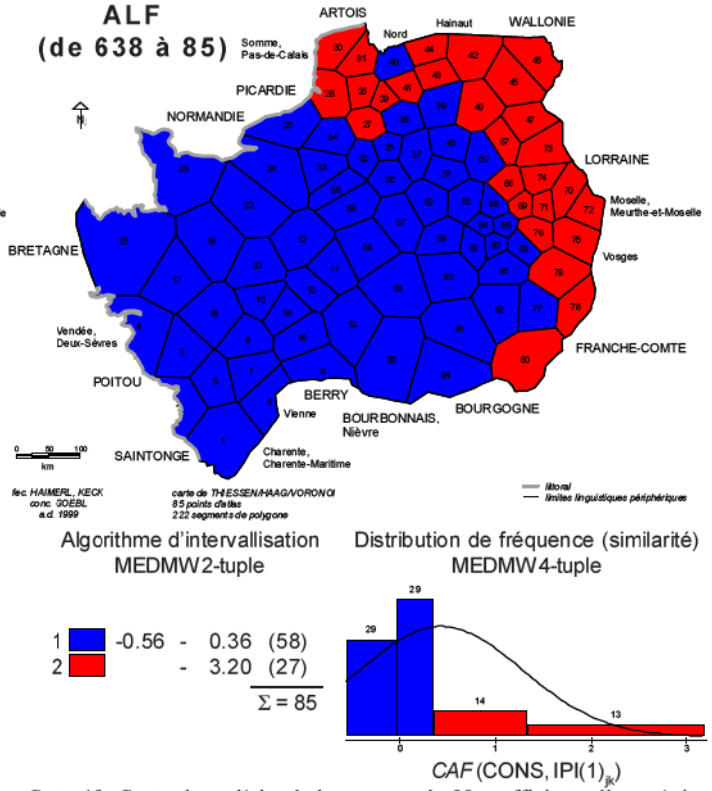
Carte 15: Carte choroplèthe de la synopse de 85 coefficients d'asymétrie de Fisher (CAF).
Corpus: Dees 1980, 170 cartes (vocalisme).
Indice de similarité: SEM_{jk}. Voir aussi le paragraphe 2.4.



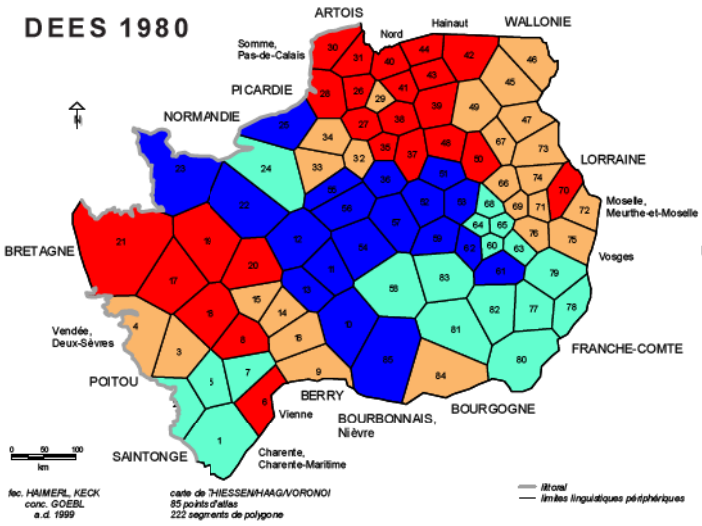
Carte 16: Carte choroplèthe de la synopse de 85 coefficients d'asymétrie de Fisher (CAF).
Corpus: ALF, 612 cartes de travail (vocalisme).
Indice de similarité: $IPI(1)_{jk}$. Voir aussi le paragraphe 2.4.

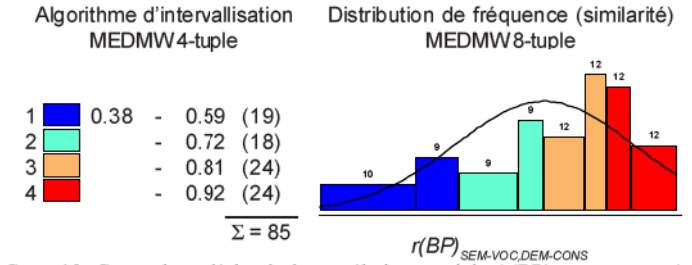


Carte 17: Carte choroplèthe de la synopse de 85 coefficients d'asymétrie de Fisher (CAF).
Corpus: Dees 1980, 78 cartes (consonantisme).
Indice de similarité: SEM_{jk}. Voir aussi le paragraphe 2.4.

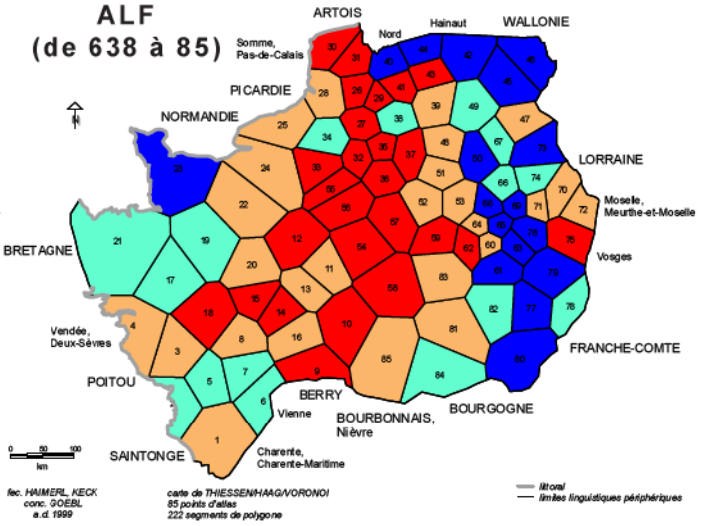


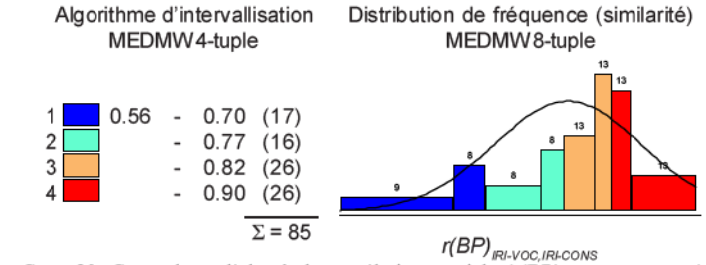
Carte 18: Carte choroplèthe de la synopse de 85 coefficients d'asymétrie de Fisher (CAF).
Corpus: ALF, 505 cartes de travail (consonantisme).
Indice de similarité: $IPI(1)_{jk}$. Voir aussi le paragraphe 2.4.



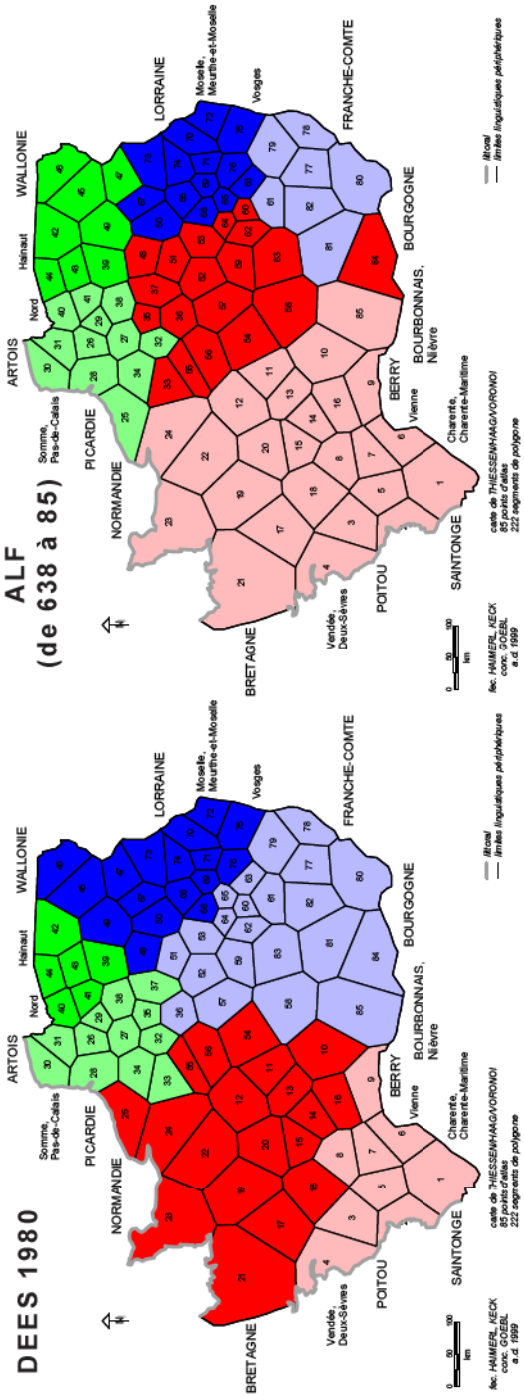


Carte 19: Carte choroplèthe de la corrélation spatiale ($r(BP)_{SEM\text{-}VOC,SEM\text{-}CONS}$) entre les aménagements vocalique et consonantique du domaine d'Oïl vers 1300.
Coefficient de corrélation: $r(BP)_{jk}$. Voir aussi le paragraphe 2.5.





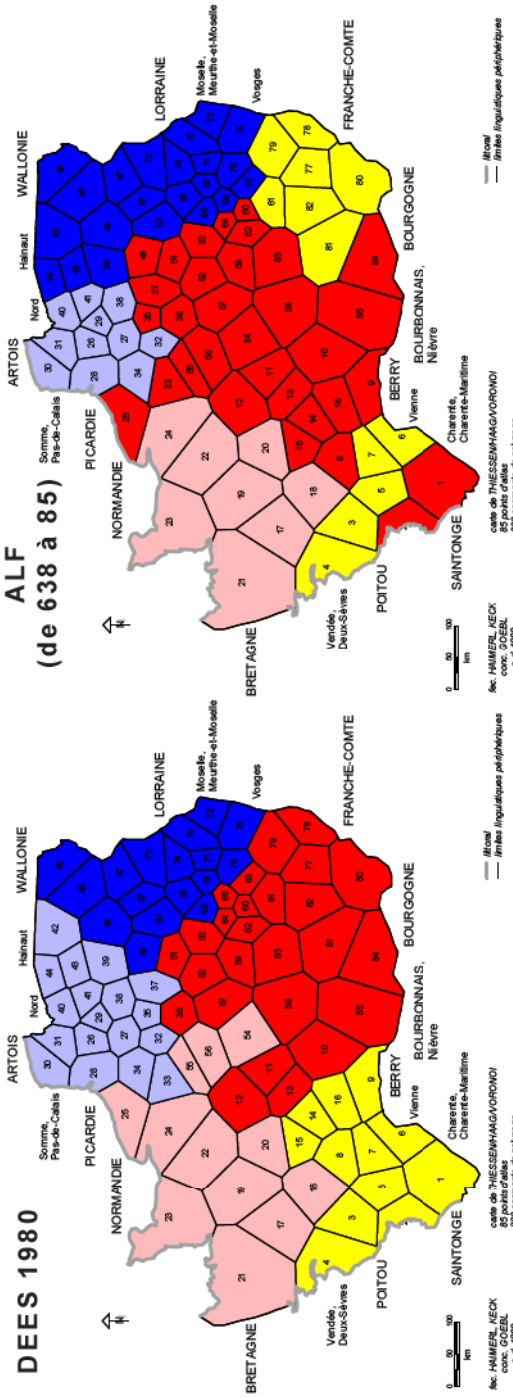
Carte 20: Carte choroplèthe de la corrélation spatiale ($r(BP)_{IRI(1)\text{-}VOC,IRI(1)\text{-}CONS}$) entre les aménagements vocalique et consonantique du domaine d'Oïl vers 1900.
Coefficient de corrélation: $r(BP)_{jk}$. Voir aussi le paragraphe 2.5.



Carte 21: Classification dendrographique de 85 points-DEES.
Algorithme agglomératif: méthode de WARD.
Corpus: Dees 1980, 268 cartes (données intégrales).
Indice de similarité: SEM_{jk}. Nombre des chorèmes et dendrèmes coloriés: 6. Voir aussi le paragraphe 2.6.

Carte 22: Classification dendrographique de 85 points-ALF.
Algorithme agglomératif: méthode de WARD.
Corpus: ALF, 1687 cartes de travail (données intégrales).
Indice de similarité: $IPI(1)_{jk}$. Nombre des chorèmes et dendrèmes coloriés: 6. Voir aussi le paragraphe 2.6.

©Universitat de Barcelona



Carte 23: Classification dendrographique de 85 points-DEES.
Algorithme agglomératif: Complete Linkage.
Corpus: Dees 1980, 268 cartes (données intégrales).
Indice de similarité: SEM_{jk}. Nombre des chorèmes et dendrèmes coloriés: 5. Voir aussi le paragraphe 2.6.

Carte 24: Classification dendrographique de 85 points-ALF.
Algorithme agglomératif: Complete Linkage.
Corpus: ALF, 1687 cartes de travail (données intégrales).
Indice de similarité: $IPI(1)_{jk}$. Nombre des chorèmes et dendrèmes coloriés: 5. Voir aussi le paragraphe 2.6.

©Universitat de Barcelona